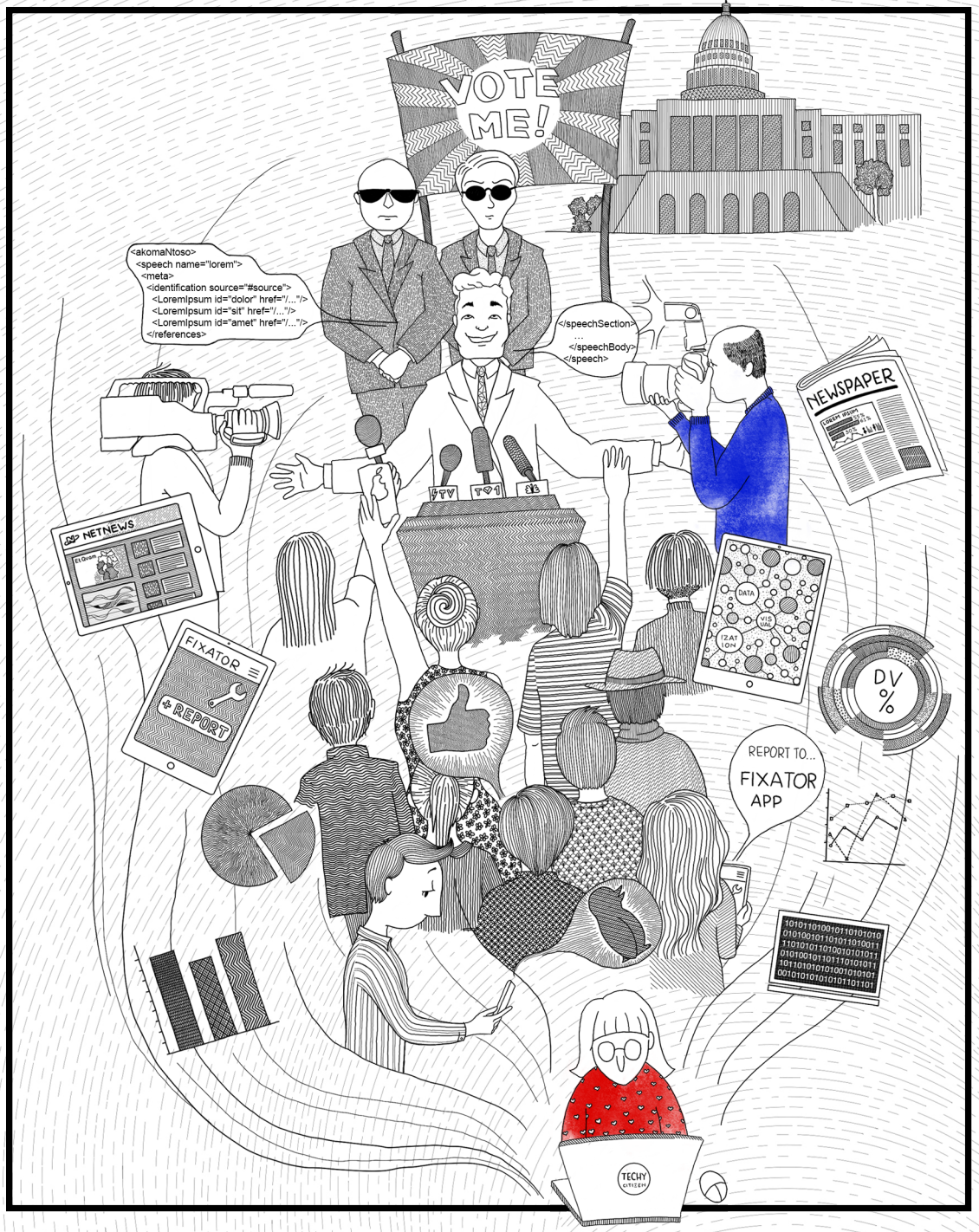
Knowing what counts

Baack, Stefan

Knowing what counts

How journalists and civic technologists use and imagine data



Stefan Baack

Knowing what counts

How journalists and civic technologists use and imagine data

Stefan Baack

Knowing what counts

Cover design: Stella Widiasanti Baack

ISBN (printed version): 978-94-034-0572-8
ISBN (digital version): 978-94-034-0571-1

university of groningen

Knowing what counts

How journalists and civic technologists use and imagine data

PhD thesis

to obtain the degree of PhD at the
University of Groningen
on the authority of the
Rector Magnificus Prof. E. Sterken
and in accordance with
the decision by the College of Deans.

This thesis will be defended in public on

Thursday 19 April 2018 at 16.15 hours

by

Stefan Baack

born on 5 November 1984
in Sande, Germany

Supervisors
Prof. M.J. Broersma
Prof. T.A.C. Witschge

Assessment Committee
Prof. B.J. Brouwers
Prof. H. Kennedy
Prof. C.W. Anderson

Acknowledgements

It is often stated that writing a dissertation can be a solitary experience. While there is some truth to that, I never felt that I was really working alone. My time as a PhD student was full of new friends, colleagues and great experiences. When I started to write these acknowledgements, I quickly realized that there are too many people I would like to thank. If that is not a sign that I was having a great time, I don't know what is. Of all the people I could mention, I want to specifically thank a few here.

Most importantly, I want to thank my wife Santi. Her love, patience, and support kept me going. Santi, I don't know how I would have been able to finish this dissertation and avoid starvation without your love and awesome cooking!

Second, I want to thank my supervisors Tamara Witschge and Marcel Broersma. I couldn't have hoped for better supervisors. They challenged me to question my ideas, gave me the freedom to experiment, and the time to reflect. Without Tamara, I'm not even sure if I would have started a PhD. She did not just support the writing and finishing of the thesis itself, but also helped me to manage living and working across two countries. Tamara, you opened many doors for me and I owe much to you. Thank you for everything!

Finally, and in no particular order, I want to thank some of my closest friends from Groningen: Alisa Van de Haar, Yu Sun, Joëlle Swart, Rik Smit, Todd Graham and Amanda Brouwers. Hugs to all of you!

Table of Contents

1. Introduction

In both academic and non-academic discourse, modern societies are said to be increasingly 'driven by data' (cf. Mayer-Schönberger and Cukier 2013; van Dijck 2014; Kitchin 2014). In lieu of relying on small, representative samples, datafication denotes a shift towards the steady quantification of social life, where social actions are continuously rendered into quantifiable data in order to enable "real-time tracking and predictive analysis" (van Dijck 2014, 198). Whether or not we take the oft-revolutionary rhetoric around datafication and big data for granted, it has become a "legitimate means to *access*, *understand* and *monitor* people's behavior" among commercial actors, governments and scientists (van Dijck 2014, 198). Despite a consensus that datafication has "far-reaching consequences to how knowledge is produced, business conducted, and governance enacted" (Kitchin 2014, 2), its implications for democratic practices are highly controversial.

On one hand, datafication has been described as a potential threat that could undermine the agency of publics. As Couldry and Powell (2014, 4) summarize, datafication might disconnect "system and experience" because the traces of data that people leave behind are often unconscious and not meaningful to them, and the insights that this gleaned data generates for companies or governments are often not "folded back into the experience of everyday life". The ways in which continuous data collection is changing how news media companies operate and are subsidized is said to create incentives to build 'filter bubbles' (Pariser 2011) which would increase fragmentation. Zuboff (2015) even suggests that the economic model that drives datafication is

fundamentally anti-democratic. Others have argued that datafication might increase social discrimination, as generating predictions based upon previous records (e.g. from crime data in predictive policing) may reproduce and potentially reinforce biases inherent upon those records (cf. Barocas and Selbst 2016).

On the other hand, datafication is also associated with a strengthening of democratic values and the empowerment of actors who aspire to work in a public interest. For example, many accounts of data journalism discuss the potential for data technologies to support journalistic autonomy by lowering the costs of investigative reporting and by creating new business models (cf. Anderson 2013b; Gray, Bounegru, and Chambers 2012; Flew et al. 2012). Beyond journalism, open data initiatives around the world are making government data freely accessible online, and promise to make "'open book' governance possible for the first time" (Margetts 2013, 167). Although these initiatives often fail to fulfill this promise (cf. Bates 2012), the concepts of open data and datafication are nevertheless strongly associated with a strengthening of democratic values, and rising levels of transparency and accountability.

Both the negative and positive accounts tend to treat processes of datafication as a self-contained and "unifying media logic (…) beyond the domain of human agency" (Hepp 2016, 927). The warnings about the negative implications often stress the way datafication is used to facilitate commercial aims or government surveillance. Critics identify commercial actors and governments as key drivers of datafication today, and shed light on the problems which may arise as a result of an uncritical embrace of data technologies. The positive accounts tend to focus on how the technological potential could help to facilitate journalistic, or more generally, democratic

practices and values. While all of these accounts provide valuable contributions, justice is not being given to the complexity of the transformations that are taking place if it is assumed that commercial actors and governments are (and always were) the *only* relevant drivers of datafication, or if we are considering the democratic potential of datafication *solely* through its technological potential.

A more nuanced approach is needed if we are to find ways to "enlist processes of datafication into the service of social progress" (Gray 2016). With this thesis, I contribute to a growing body of research which suggests that we should study how people use data "to meet their own ends (…) for broader social, civic, cultural or political goals" (Couldry 2014, 892). Rather than abstracting the democratic potential of datafication by looking at its technological features, I consider how datafication is appropriated and advanced by actors who aspire to work in a public interest, and how this might "shift agency and representation" (Couldry and Powell 2014, 4). My premise is that the ways in which processes of datafication "*sustain, undermine and transform vital public values*" (Kennedy, Poell, and van Dijck 2015, 2) very much depend on how these processes are facilitated, and by whom.

Studying datafication through pioneer communities

To develop a more nuanced approach to the study of how processes of datafication affect public values, I study 'pioneer communities' who use and appropriate data technologies in novel ways. According to Hepp (2016, 928), pioneer communities are influential intermediaries "between the development and the appropriation of new media technologies". They are among the first

to appropriate new technologies and develop "a horizon of *possibility*" (Hepp 2016, 919), i.e. they imagine and exemplify how a technology may be used, and how it might change the status quo. The discourses and practices that pioneer communities generate provide orientation for others, and influence the wider adoption of a technology. A famous example of a pioneer community that is related to datafication is the quantified-self movement, which has been a pioneer in developing techniques and visions for self-measuring (Hepp 2016; Nafus and Sherman 2014). As Hepp (2016, 928) explains, the study of pioneer communities allows us to link broad transformations like datafication to the practices and visions of different stakeholders in these processes, and to treat social transformations as something that is imagined and promoted by specific groups.

To study how datafication is appropriated by actors who aspire to work in a public interest, I examine two pioneer communities from the field of journalism and civil society: data journalists, which I use here as an umbrella term for journalists who are engaged in quantitative forms of journalism (cf. Anderson 2015), and activists in the open data or civic tech movement, which I will abbreviate as data activists (cf. Schrock 2016; Milan and Van der Velden 2016).

Data activists develop projects that attempt to make engagement with authorities easier for citizens. This includes the development of problem reporting websites which make it easier to report local infrastructure issues to local government. Other examples of initiatives launched by data activists include parliamentary monitoring websites, freedom of information (FOI) websites designed to help users to submit freedom of information requests to public institutions, and additional websites which are designed to help users to identify and contact their representatives in parliament. In its modern form,

this type of activism is the result of a convergence between "communities of technological and political openness" (Yu and Robinson 2012, 195). Originally powered by volunteers in their spare time (cf. Townend 2008), the civic tech sector has now evolved into a global phenomenon that is being embraced by governments, corporations and foundations.[1]

'Data journalism' is the label that is now commonly used to describe all forms of journalism that work with quantitative data. Quantitative forms of journalism are not new (Anderson 2015), but the affordances of the new media environment, combined with datafication, have resulted in an "extended dimensioning and accessibility of computational opportunities inside and outside of news organizations" (Gynnild 2014, 718). What makes data journalism methodologically and epistemologically different from previous forms of quantitative journalism is disputed, but it is clear that data journalism is responding to datafication, and utilizes quantitative techniques on a much larger scale (see Chapter 2). Journalism is said to be taking a "quantitative turn" (Petre 2013; cf. Coddington 2015), at least in the US and Europe, where research has shown how news media organizations are increasingly incorporating quantitative practices and building dedicated data journalism teams, albeit in highly stratified and uneven ways (Fink and Anderson 2015).

[1] This is evidenced by numerous of international conferences and collaborations in this sector. For example, the 'Open Government Partnership' is a multilateral initiative founded in 2011 with 75 participating countries by 2017 (https://www.opengovpartnership.org/about/about-ogp). 'Code for All' is an example of an international network of civic tech and open data organizations with members in 13 different countries as of 2017 (https://codefor-all.org/partners/).

I focus specifically on data activists and data journalists for two reasons:

1. Data journalists and data activists are pioneers for the use of data in the field of journalism and civil society

Journalism and civil society generally play a key role in assembling publics (cf. Anderson 2013a, 175). The ways in which they use data to produce knowledge and how their reliance on data shapes the images of the publics they aim to assemble inevitably affects the collectives we form – and there are several indications that data activists and data journalists shape the use of data in these fields. Data journalism was introduced and made popular in the mid-2000s by exceptionally large and successful newsrooms such as The New York Times and The Guardian, who continue to do "truly pioneering, even revolutionary, computational journalistic work" (Fink and Anderson 2015, 479). The adoption of quantitative techniques by other news media was also influenced by dedicated journalists such as Adrian Holovaty (2006) who became famous for his much-cited manifesto, *A fundamental way newspaper sites need to change* (cf. Gynnild 2014, 721). Holovaty, along with other journalists who possessed a sense of mission typical for pioneer communities (Hepp 2016, 925), helped to shape the discourse around data-driven and computational techniques in journalism via blogs, conference presentations and other platforms (cf. Parasie and Dagiral 2013).

Data activists, similarly, acted as pioneers for the use of data within civil society. By making information accessible online, and then turning this information into structured data with which to build new services, data activists combined information access with open source values such as collaboration and sharing. This combination of accessibility and reusability brought legal and technical concepts of openness together in new ways (Yu

and Robinson 2012, cf. Chapter 2). In addition, some data activists organized in civic tech organizations self-identify as pioneers, and train other civil society or media organizations to work with data or to use the tools provided by data activists to their advantage. An example of this is the 'School of Data' organized by Open Knowledge, which is active in various countries. Non-profit civic tech organizations, such as the Sunlight Foundation in the US or mySociety in the UK, were also among the first organizations to advocate for open data policies in the US and Europe (Schrock 2016).

In short, data journalists and data activists have pro-actively helped to facilitate the use quantitative techniques in key areas of public space. Given the potential influence of pioneer communities, studying data journalists and data activists provides an opportunity to understand how datafication is appropriated by journalists and civil society actors more broadly.

However, while the current research on these pioneer communities has paid a lot of attention to data journalism (cf. Fink and Anderson 2015, 476), little is known about the practices, values and self-understandings of data activists. The majority of existing research about data activists has been conducted by civic tech organizations and their funders, who are primarily interested in the impact of civic tech applications, such as who is using civic tech applications, how the applications are being used, and why (cf. Escher 2011). Researchers from media and journalism studies often look at civic tech or open data from the perspective of political economy. This work tends to be highly critical, and often describes open data or civic tech as expressions of existing power structures that are "empowering the empowered" (Gurstein 2011). For example, Bates (2012) argues that the open data movement uncritically embraces the open data initiative of the British government, which utilizes open data to follow a neoliberal agenda that aims for further deregulation, and

the outsourcing of public services to private actors. For Gregg (2015), civic hackathons are connected to politics of austerity, and serve to normalize sacrificial working conditions.

While such critique is important and necessary, it appears to dismiss activism around open data and civic tech as strictly problematic while leaving out further analysis. As Kennedy (2016, 217) points out, we can argue that activists in the open data or civic tech movements do play a role in sustaining existing power structures but it would be "empirically inaccurate to suggest that they *only* do this". As described above, open data and civic tech have become global phenomena that pioneered the use of data to support forms of civic engagement and activism. Because the practices and visions developed by such pioneer communities provide orientation for others, analyzing them is important regardless of whether we ultimately view them as problematic or not.

2. The entanglements between data journalists and data activists have broader implications for journalism and civil society actors

The second reason for focusing on data activists and data journalists is that these actors are increasingly entangled, and appear to complement each other with ease. This relationship can be attributed to parallel developments around open data and within journalism over the past decade. First, technologists engaged in the open data and civic tech movements have been attracted by the idea that their computational skills can be used to support a public good, and critical 'watchdog journalism' in particular (Parasie and Dagiral 2013, 861; Lewis and Usher 2013). Second, news media companies began hiring developers in the mid-2000s with the hope of increased traffic (e.g. through interactive visualizations), new business models, and (re)gaining public trust

by establishing themselves as "data-custodians of the true and the quantitative" (Nussbaum 2009; cf. Parasie and Dagiral 2013). An early example of this is the 'Interactive News Technology department' of The New York Times (Royal 2010). Third, both of these developments were reinforced by US foundations, which started initiatives to bring technologists into the newsrooms. Chiefly important here is the Knight Foundation, which transformed from a journalism-centric organization to a "boundary-spanning agent" (Lewis 2012a, 329) that has connected with other professions and non-profit foundations. This is typified by the 'Knight News Challenge', which invites ideas from groups both inside and outside of the field of journalism to better serve the "information needs of communities" (Lewis 2012a).

Some news media organizations have been trying to take advantage of the developments outlined above by re-interpreting the unique ways in which developers and programmers think about technology and data "into the language of news" (Lewis and Usher 2013, 604). Journalists have adopted some of the activists' practices and ideas, used some of their applications for their own investigations, and occasionally engaged in direct cooperation with data activists. At the same time, civic tech organizations have reached out to journalists to use their existing tools in various ways (cf. Townend 2008), have sought collaboration with media organizations, and developed customized tools especially for journalists. It is therefore unsurprising that practitioners and foundations often lump civic tech and data journalism together, and focus on their similarities (cf. Howard 2014c; Townend 2009).[2]

[2] For example, the 'Code for Africa' initiative describes itself as "Africa's largest Civic Technology + Data Journalism Initiative" and is supported by partners active in both fields,

The mutual awareness and cooperation between data journalists and data activists indicate that although these actors come from culturally, ideologically, and institutionally distinct traditions, their shared reliance on data and their common orientation towards the public has made their work similar, or at least compatible. Given the status of data activists and data journalists as pioneer communities, i.e. as communities whose discourses and practices provide orientation for others, these entanglements are significant, as they may influence the relationship between journalism and civil society more broadly. Yet, the current research literature tends to either look at data journalists and data activists in isolation, or to narrowly focus on the direct interactions between them.

For researchers in media and journalism studies who are interested in data activism, the relationship between data activists and data journalism has not been a primary focus thus far. Most of the current research on data journalism has focused on how it is being integrated in newsrooms (Fink and Anderson 2015). The ways in which data journalists and actors from the technology sector are mutually influencing each other have received far less attention from the research community, despite the acknowledgement that the news-making-process involves a great "diversity of actors, discourses and relationships" that influence how news is found, produced and circulated (Domingo, Masip, and Costera Meijer 2015, 53). While there is a growing body of research on the relationship between journalists and technologists rooted in open source culture (see Chapter 2), this research tends to primarily focus on direct

like the International Center for Journalists or the Omidyar Network. See https://codeforafrica.org/.

interactions between these actors in joint workshops or within newsrooms. Moreover, researchers usually ask how this interaction influences journalism, not whether and how journalists influence others.

Analysis of the direct interactions between data activists and data journalists is important, but if we are to fully understand their dynamics and subsequent implications for journalism and civil society, we must search more broadly. What is generally overlooked are the consequences as a result of the long-term co-existence of non-profit civic tech organizations and news media with dedicated data journalism teams. How does the continued exchange and awareness of each other affect how the actors involved work and understand their work, and how does this influence their use of data? We need to understand the nature of their entanglements in order to more fully grasp how their relationship affects the way these actors appropriate and expand datafication.

A focus on practices

To help rectify the above shortcomings of the current research literature regarding data activists and the entanglements between data activists and data journalists, I critically examine data journalists and data activists as pioneer communities following a practice theory approach. A focus on practices is valuable because pioneer communities do not just communicate new ideas; they *perform* them. Pioneer communities become influential because they act as exemplars. Their daily practices are expressions of their broader visions, and we must consider them together if we are to understand their influence. This link between practice and discourse has been illustrated by Taylor (2004), Schatzki (2001), and Couldry (2004, 2012). As these authors have shown, practices are not merely material actions but complex bundles of actions which

embody "modes of understanding" (Taylor 2004, 31). For Schatzki (2001), practices are teleological nexuses of doings and sayings in the sense that a practice is always performed with a certain goal in mind, which more quotidian actions should help to achieve. Practices are organized hierarchically: at the top is a person's end, an activity that does not help compose any further activity (Schatzki 2001, 15). For example, a journalist's end might be the rather abstract idea of holding politicians accountable. Various day-to-day actions are then carried out for the sake of this end, such as interviewing relevant subjects, writing, reporting on particular issues, etc.

More fundamentally, Taylor (2004) suggests that practices are not simply goal-oriented, but intrinsically tied to much broader 'social imaginaries'. Social imaginaries describe the ways in which groups of people 'imagine' their social surroundings through images and stories, a common understanding that makes possible "common practices and a widely shared sense of legitimacy" (Taylor 2004, 23). In contrast to theories about the social world or norm awareness, social imaginaries describe a complex 'background' understanding that people call upon to make sense of what they are doing. Taylor (2004, 25) claims that every collective has a specific 'repertory' of actions at their disposal. The practices of actors who share a social imaginary "reflect a commitment to working out (…) shared concepts" (Kelty 2008, 42).

Changes in our social imaginaries are either expressed through changes in practices or through changes in the meanings and understandings that underlie existing practices (Taylor 2004, 30). Similar to Hepp (2016), Taylor (2004, 30) argues that such transformations usually take place when "certain groups and strata of the population" develop new imaginaries and practices which "recruit a larger and larger base". Taylor's (2004) work shows that the 'horizon of possibility' that pioneer communities provide for others is comprised of both

material actions and the broader visions that those actions should bring to fruition. To understand the role of data journalists and data activists as pioneer communities, we thus need to consider how they work with data, what broader imaginaries underlie their use of data, and what they strive to accomplish through these practices.

The second reason to focus on the practices of data journalists and data activists is that practice-focused research paradigms can be both open-minded and sensitive to power structures. I specifically draw from Couldry (2004, 2012), who relies on Schatzki's (2001) definition of practices. Couldry (2004, 117) aims at developing more nuanced understandings of human agency by focusing on the "open-ended range of practices". At the same time, he adds an emphasis on power relationships to Schatzki's (2001) conceptualization of practices. For Couldry (2012, 34), practices "are not bundles of individual idiosyncrasies; they are social constructions that carry with them a whole world of capacities, constraints and power".

This dual emphasis on openness and structure is helpful because it captures the tension that is inherent to the study of pioneer communities. By definition, pioneer communities necessarily transform existing practices; this complicates the process of putting them into well-established categories. This classification paradox also applies to data journalists and data activists, both of whom are unstable and unsettled in their practices and state of institutionalization, possess a loose terminology with no common and clear definitions, and are composed of a diverse set of actors (see Chapter 2). It is therefore difficult study the entanglements between them with the use of predefined conceptions and delineations. At the same time, their unsettledness does not mean that there is no structure. Given the *anticipated* influence of pioneer communities, we must study these communities in ways that help to explore the power

relationships embedded within their social imaginaries. This is especially important for actors such as data journalists and data activists, who attempt to assemble particular visions of publics and affect how we make sense of our social surroundings. Gaining an understanding of the social imaginaries now will help us to understand the implications of their practices later, once they have become more widespread.

Research questions and case study design

Due to limitations of the current research literature and the focus on practices outlined above, this thesis addresses two research questions designed to critically examine data journalists and data activists as pioneer communities:

1. What is the role of data in the social imaginaries and practices of data activists and data journalists?

2. How do the practices and imaginaries of these actors diverge and converge, and how does this shape the entanglements between them?

The first question asks how data journalists and data activists understand their work, the public services they aim to provide, and what role they attribute to (their conceptualizations of) data. How do they utilize data, and what role does data play in their broader imaginaries? Finding answers to these questions provides a basis to answer the second research question, in which I examine the entanglements between data journalists and data activists and theorize about the resulting implications.

To answer these research questions I conducted three case studies, which mutually informed each other. The majority of the empirical work has been focused on data activists. Compared to data journalists, little is known about

the practices and imaginaries of data activists. Therefore, more fundamental and exploratory work was necessary to examine their role as a pioneer community, as well as their entanglements with journalism. The first two case studies exclusively deal with data activists, and focus on 'best practice' organizations. A focus on best practice organizations is a useful starting point to explore fields that have not been thoroughly researched, because these organizations provide orientation for others as exemplars (see Chapter 3 for more details).

First case study: Datafication and empowerment. How the open data movement re-articulates notions of democracy, participation, and journalism.[3]

It seems clear that open data activists advocate for the legal and technical openness of government data (cf. Davies 2010); but what do they aim to realize *through* the opening of government data? The first case study (Chapter 4) critically examines how open data activists imagine the publics that they aim to assemble, and what the implications are for the agency of datafied publics (cf. Couldry and Powell 2014, 4; Kennedy and Moss 2015). As argued above, to understand how processes of datafication sustain, undermine, or change public values, we must understand how such processes shape the assemblage of publics.

To address these questions this case study focuses on the Open Knowledge Foundation Germany (OKF DE), a non-profit organization founded in 2011, and the most visible and relevant actor in the German open data movement. I show how the OKF DE applies practices and values from open source

[3] Published in *Big Data & Society*, see Baack (2015a).

culture to data. Using a model developed by Kelty (2008) to analyze how practices and values from open source culture are applied to new domains, I show how members of the OKF DE take some of the key practices of open source (such as sharing source code or coordinating collaborations) and apply them to the creation, use, and analysis of data in order to change the relationship between governments and their publics. This bringing together of open source culture and data leads the OKF DE to develop a modulated version of open source governance that fundamentally relies on the availability and modifiability of data, which raises interesting questions about the agency of datafied publics.

Second case study: Civic tech at mySociety. How the affordances of data shape data activism[4]

The second case study (Chapter 5) expands upon the first one by placing emphasis on how activists use data to "meet their own ends" (Couldry 2014, 892), i.e. how they utilize data to realize their imaginaries and how they position themselves within the public arena. Describing their data-related practices and the underlying imaginaries in-depth is important for understanding the potential impact of data activists as pioneers, as this knowledge will then enable future studies to examine if and how other actors may have adapted these practices. I draw from Nagy and Neff's (2015) concept of imagined affordances to explore how data is used, and how it is thought to contribute to the realization of the visions of mySociety, a non-profit organization from the UK. mySociety is one of the oldest and most successful

[4] Published in *Krisis* (http://krisis.eu/), see Baack (2018).

civic tech organizations to date, and has pioneered many of the civic tech applications that have later been replicated by groups in other countries.

The concept of imagined affordances insists that what a particular technology affords is not merely a question of its physical properties or functionality, but also how users and designers imagine what a technology is for. I thus examine how members of mySociety rationalize and utilize data in specific ways to support their agenda. The study shows how mySociety imagines ways in which data expands the agency of publics towards governments: ways of using data that would enable citizens to better influence and interact with governments or other powerful institutions. mySociety is trying to facilitate civic engagement and, by extension, create a more participatory culture. Structured data is used to remove frictions that make civic engagement and monitoring governments more difficult and time-consuming for citizens. The study further demonstrates that the "political and democratic possibilities of data" (Milan and Van der Velden 2016, 8) are both culturally and historically situated and cannot be subsumed under unifying media logics (see above).

Third case study: Practically engaged. The entanglements between data journalism and civic tech[5]

The first two case studies have formed the basis for the third and final one, which explores the entanglements between data activists and data journalists (Chapter 6). The third case study, in contrast, does not focus on any specific best practice organization. Due to the lack of research on the practices and imaginaries of data activists, a focus on influential organizations was useful in

[5] Published in *Digital Journalism*, see Baack (2017).

the first two cases. For data journalism, however, I relied on the rich research literature exploring how data journalism is integrated in newsrooms, and how data journalists understand their work (see Chapter 2). To further explore how data journalists are entangled with data activists, I sought a broad range of viewpoints and interviewed data journalists in diverse organizational settings such as national and local news media, startups, and non-profit newsrooms. This allowed me to complement the data that I collected in the previous case studies, which was included in the final analysis to directly compare the perspectives of data activists and data journalists.

In this study, I critically examine how data journalists and data activists interact with each other, how similar or different their ambitions are, and how they complement each other. Additionally, I ask what the consequences of their entanglements are for producing knowledge in the public interest. How do data activists, data journalists in different organizational settings, or individuals who are actively engaged in both data journalism and data activism understand their own work, how do they view each other, and how do they position themselves professionally?

I show how journalists and activists together form what can be called a community of practice (Wenger 1998) or a figuration, i.e. a network of actors that are linked through interlocking practices and shared meanings (Couldry and Hepp 2017, Chapter 4). Along the axes of 'facilitating' and 'gatekeeping', I identify four different groups that are closely connected yet distinct in their self-understanding, how they position themselves professionally, and how they use data. Actors who are interested in facilitating seek to enable the users of their tools to take action themselves, while gatekeeping seeks to direct public debate by identifying and presenting publicly relevant information. Practices of facilitating and gatekeeping have traditionally been understood as

competing 'logics' (Lewis 2012b). My findings challenge this assumption and show how the ongoing datafication of social life allows for practices of facilitating and gatekeeping to exist along a shared continuum, and to mutually reinforce each other.

The findings of my research demonstrate that we should not limit ourselves to the study of how reliance on data changes individual actors or fields such as journalism; we must also ask how datafication is connected to emerging figurations, and what the implications of those figurations are if we are to understand how datafication affects public knowledge production and the assemblage of publics.

Thesis structure

In Chapter 2, I examine the existing research literature, focusing on three aspects relevant to understanding data journalists and data activists as related pioneer communities. First, I look into the historical trajectories of these practices. This exploration is necessary in order to understand how they became 'pioneers' and how they continue, or break certain traditions. Second, I look into research examining the relationship and direct interactions between data journalists and data activists, or more broadly, actors from the larger technology sector who are rooted in open source culture and interested in journalism (programmers, computer scientists and so forth). Third, I discuss how both data journalists and data activists are connected to notions of monitorial democracy. I review how this concept has been applied to theorize these actors, and discuss the implications for this study.

Chapter 3 presents the methodological framework based on the focus on practices outlined above. I combine grounded theory's principle of theoretical

sampling, in which an initial data collection is continuously expanded to elaborate and refine the emerging theory, with a qualitative multi-methodological approach, relying on interviews, participatory mapping, content analysis, ethnography, and Digital Methods. This methodological approach allows me to be sensitive to the open-endedness of practices and sense-making processes which take place among data journalists and data activists. These methods also provide an insight into how data activists and data journalists connect their broader ambitions and values to certain practices around data, i.e. with the ways in which data is being gathered, analyzed, shared, and presented.

Chapters 4-6 present the findings of the three case studies outlined above. Chapter 7 summarizes the overall results of the three articles, and discusses their broader implications. By showing that there are actors who advance datafication for the purpose of serving a public interest, this thesis contributes to our understanding of datafication as a contested process, and demonstrates the value of studying the current state in flux qualitatively.

2. Theoretical context

In this chapter, I examine the existing research on three aspects relevant to understanding data journalism and data activism as pioneer communities: 1) their history and predecessors, 2) the relationship and direct interaction between data journalists and data activists, and 3) the implications of these phenomena, i.e. their potential causes and effects.

To understand the role of data journalism and data activism as pioneer communities, and to assess the implications of the practices and imaginaries they develop, we have to ask what exactly makes data journalists and data activists 'pioneers'. How do they continue or break off from earlier traditions? Data journalism and data activism are not spontaneous reactions to new technological affordances: they are culturally and historically situated practices. How data is understood, utilized, and connected to ideas about democratic publics has changed over time, and these historical trajectories influence the ways in which data is being used today. Anderson (2015) illustrates this in his study about the history of quantitative practices in journalism. He shows that this is not a history of continuity, i.e. "of ever more sophisticated quantitative work", but of "transformation and rupture (…) less inevitable and timeless than it is the product of deliberate choices intersecting with historical structures" (Anderson 2015, 351–52).

In the first two sections, therefore, I examine research literature that looks at the historical roots of data journalism and data activism. I show how data activism is rooted in movements for technological and political openness (specifically the freedom of information and open data movements), and how the discourse around the term 'civic tech' reflects the similarities and

differences between these roots. For quantitative journalism, I explore the similarities and differences of three labels which are commonly used to refer to quantitative practices in journalism: computer-assisted reporting, data journalism, and computational journalism. Comparing the history and use of these terms reveals shifts in how the role of data in journalism is understood.

In the third section, I look at research relevant for understanding the relationship between these groups. Because there is no research that specifically investigates the entanglements between data activists and data journalists, I look at three related issues that have received attention: 1) the interactions and collaborations between journalists and actors from the larger technology sector (e.g. at events or in joint projects), 2) how open source culture is compatible, or clashes with, professional journalism, and 3) the spread and modulation of open source culture.

In the fourth section, I discuss the possible implications of data activism and data journalism. I show how their practices and self-understandings resonate with the concept of 'monitorial democracy', developed by Schudson (1998, 2015) and Keane (2009). I further discuss how these concepts have been applied to data activism and data journalism, and consider the implications for methodological design and the empirical analysis.

Historical roots of data activism

As mentioned in Chapter 1, in its modern form, the open data and civic tech movement is a convergence between "communities of technological and political openness" (Yu and Robinson 2012, 195). In this section I examine this convergence in more detail, and what the implications are for how open data activists and civic technologists act as pioneer communities. First, I

discuss how the open data and civic tech movement consists of groups with different notions of 'openness'; these notions of openness are compatible, but can be difficult to align.

The second part of the following section looks more closely at 'civic tech' as an emerging umbrella term. The emergence of this term, and the practitioners' struggle to define it, reflects an effort to reconcile the diverse roots of the communities they form. The struggle to develop common visions and identities shapes how activists understand their work and come to develop a 'sense of mission' as a pioneer community. As both the research literature and the discourses among activists demonstrate, the horizon of possibilities civic technologists and open data activists develop is still an evolving one.

Between legal and technical openness

What I describe as the open data and civic tech movement in this thesis is a blend of diverse interest groups with different historical backgrounds. This diversity is reflected in the main arguments commonly used to justify the opening of government data. Janssen (2012) summarizes these arguments as follows: First, open data is considered vital for transparency and accountability, because in order to hold the government accountable one must be aware of governmental activities. Second, open data is argued to be a driver of participatory governance because it allows citizens to be informed, and creates opportunities to build new platforms with lower thresholds to participate. Third, open data would stir innovation and economic growth, as new applications and services could be made based on this data. Fourth, open data would make public services more efficient because they would be able to access data held by other authorities, and receive more feedback from the

public. Through these four justifications, we can roughly separate the legal and technical notions of openness.

Legal openness is advocated by what Janssen (2012) calls the right-to-information (RTI) movement. This movement emerged after World War 2 in the US, when journalists lobbied for the right to access government information. In 1966, the 'Freedom of Information Act' (FOI) turned their demands into a legal framework by giving every citizen the right to request previously undisclosed government information – though many exceptions remain that allow the US government to withhold information (Schudson 2015, Chapter 2). Since the 1960s, similar legal frameworks have been adopted in more than 90 countries, and the idea that citizens should have a right to information gained support from international organizations such as the United Nations (Janssen 2012). The RTI movement played a major role in the international recognition of the right to access information. It consists mainly of civil society organizations who advocate for such legal right to information frameworks to promote government accountability and public participation (cf. Janssen 2012; Yu and Robinson 2012, 186).

More technical notions of openness are advocated primarily by two groups. First are technologists with roots in open source culture. Open source stands for a mode of organization that is based on voluntary participation (Weber 2004, 62) and collaboration, inviting and incorporating contributions from potentially everyone.[6] Similar technical notions of openness are also

[6] Strictly speaking, 'free software' and 'open source' are different moral genres based on the same set of practices (Coleman and Golub 2008). Free software emphasizes social and cultural values ('free as in speech'), while 'open source' emphasizes the practical advantages for

advocated, secondly, by the profit-oriented public sector information (PSI) re-use industry. PSI primarily sells datasets and information products to public and private organizations, and is interested in 'unrefined' datasets from public bodies to create value-adding services (Bates 2012).

Since the 1990s, these movements for legal and technological openness have increasingly merged (Yu and Robinson 2012; Schrock 2016). On the one hand, RTI activists started to view the potential of the internet for increasing accessibility "as a natural extension of [the] freedom of information movement" (Schrock 2016, 587). On the other hand, a growing community of developers emerged in the early 2000s, who were interested in developing tools that would change the ways citizens engage with their governments online (cf. Townend 2008).

According to Yu and Robinson (2012), the first major project that can be described as an open government data project was OpenSecrets.org in 1998. It not only provided access to campaign finance disclosures, but also made this data machine-readable and accessible to combine "government data with third-party innovation" (Yu and Robinson 2012, 192). In 2003, volunteers in the UK developed FaxYourMP, which allowed citizens to write messages to their representatives in parliament. This idea was compatible to both RTI's notions of accountability and open source's emphasis on accessibility and

---

developing software and is more compatible with commercial interests (Kelty 2008; Coleman 2016). I use 'open source culture' to address both strands because this study does not aim at exploring hacking genres. Moreover, the actual practices behind free software and open source very much overlap and putting them together makes sense given the methodological focus inspired by practice theory in this study (see Chapter 3).

collaboration. GovTrack.us, developed by then-graduate student Joshua Tauberer in 2004, represented a "landmark in the convergence of open government [closely associated with FOI] and open data" (Yu and Robinson 2012, 192). Tauberer scraped information provided by the US Congress, and made it machine-readable and freely available through an application programming interface (API) to allow others to build their own applications based on it.

Another important step for bringing together notions of legal and technical openness were the 'Eight principles of open government data' (OpenGovData.org 2007), developed at a meeting organized by the Sunlight Foundation, a US-based non-profit organization which aims to apply "right to information principles to the Internet (…) to improve access to information about elected officials" (Schrock 2016, 587). Still, the eight proposed principles were leaning towards open source culture and defined the properties of data that would allow sharing and collaboration (e.g. machine-readability, completeness, timeliness, free licenses), instead of conditions that would guarantee government accountability (Yu and Robinson 2012, 196; Schrock 2016, 588–89). The Obama administration later adopted the notion of openness developed in the eight principles into their open government initiative (Yu and Robinson 2012).

This merging of different notions of openness has created an "ambiguity" (Yu and Robinson 2012). The 'openness' that activists in the open data and civic tech movements advocate for can refer to political accountability, better collaboration and participation, efficiency, potential business opportunities around freely available government data, or a combination of these aspects. There has been some discussion as to whether this ambiguity of 'openness' is a "blessing or a curse" (Janssen 2012), from the perspective of RTI and the

improvement of political accountability (Bates 2012; Yu and Robinson 2012). On the one hand, movements that advocate for more technical notions of openness (particularly open data advocates) and actors in the RTI movement both want more accessibility to the information held by public authorities. They both want a more 'open' political process, i.e. a more transparent and participatory political process. The combination of different notions of openness, thus, might be mutually beneficial; e.g. when governments who are interested in open data for economic reasons become more transparent (cf. Yu and Robinson 2012, 116).

On the other hand, as Janssen (2012) describes, different nuances and priorities seem to have grown throughout the development of RTI and open data. Most importantly, greater access to government data does not necessarily lead to greater political accountability and public participation. Consequentially, there are differences in the kind of data that is considered to be interesting: while technology-driven groups want 'raw', machine-readable data, RTI groups put an emphasis on "intellectual accessibility" (Janssen 2012) and are interested in any type of qualitative or quantitative information that increases government accountability. The blend of rights-based freedom of information movements, technologists with roots in open source culture, and PSI-reuse industries continues to shape activism in the open data and civic tech movement. This becomes visible in the struggle of civic technologists to define themselves and their work around the emergence of 'civic tech' as an umbrella term that is supposed to capture every aspect of this movement.

Civic Tech: Ambiguity as a virtue and a problem

Civic tech appears to be derived from 'civic hacking', a term that emerged in the early 2000s in the UK. Influential was an article on openDemocracy.net by

James Crabtree (2003), at that time a policy analyst and trustee of the 'UK Citizens Online Democracy' (UKCOD) charity group, which explored the democratic potential of the internet. Inspired by volunteer experiments like FaxYourMP (see above), Crabtree (2003) criticized the e-government initiative of the UK government for concentrating too much on making existing government services accessible online, rather than using the new technological affordances to rethink them. Using the success of Napster as an example, Crabtree (2003) argued that the UK government should instead support the development of applications that connect "ordinary people with other ordinary people", and help them to "overcome life problems". He suggested that this should be a "new agenda for e-democracy", which he termed 'civic hacking' (Crabtree 2003).

Crabtree's then-flat mate Tom Steinberg, also a policy analyst at that time, was inspired by Crabtree's (2003) ideas and wanted to "unite UKCOD with the grassroots talent at FaxYourMP" (Donoghue 2008). Steinberg initiated a 'civic hacking fund' to collect ideas on how a 'Napster for civil life' (BBC 2003) might look, and recruited many of the volunteers that helped to develop FaxYourMP, and other similar projects (Donoghue 2008). Shortly after its start, the idea to fund others to develop applications was abandoned. Instead, Steinberg's fund was turned into the first 'civic hacking organization' that developed its own applications: mySociety. In its origins, civic hacking was, thus, a mix between grass-roots technologists interested in using web-technology to change democratic processes, and policy analysts interested in a new approach towards e-government.

'Civic tech' appears to have largely replaced 'civic hacking' due to the negative connotation of 'hacking'. As a term, *civic tech* is rather generic and open to interpretation. This allows the phrase to attract a broad range of stakeholders:

governments, companies, non-profits, and transparency and open source advocates (cf. Shaw 2016). Even though there is no commonly accepted definition, civic tech is used to label every kind of project that somehow improves accessibility to public services or increases transparency: projects like the aforementioned GovTrack.us, parliamentary monitoring websites, problem-reporting websites, freedom of information websites, and more. In this way, the term *civic tech* pulls together the diverse groups described above, and partially obscures their differences.

While this obscuring has helped RTI activists, open source technologists, and PSI-re-users to merge in the first place, it is perceived as problematic by some, and has been the cause of many self-reflexive discussions about a common identity. This discourse has produced numerous broad definitions of civic tech: "the nexus of technology, civic innovation, open government and resident engagement" (Knight Foundation 2013); "the use of technology for the public good" (Stempeck 2016); "any tool or process that people as individuals or groups may use to affect the public arena" (Sifry 2014); "improving methods of getting public input into government" (Shaw 2016); "people working together quickly and creatively to help improve government" (Levitas 2013); and "serving people's need to obtain and deploy power" (Steinberg 2013). A report by the Omidyar Network, one of the biggest funders of non-profit civic tech organizations, unsurprisingly suggests that the 'civic tech sector' has common themes but lacks "a coherent and clearly articulated vision and sense of shared identity" (Donohue 2016).

These discussions suggest that civic tech is still an emerging field that struggles to develop coherent visions. This struggle might also weaken the role of civic technologists as a pioneer community, whose discourse provides orientation for others. However, such assertions are problematic, as we know little about

the different types of actors involved in civic tech. It is worth re-emphasizing that this thesis focuses on non-profit organizations (OKF DE and mySociety, see Chapter 1) because it is interested in how datafication affects public values. Uncertainty and ambiguity within the larger civic tech sector does not mean that individual organizations within this sector do not possess a clear vision and sense of mission typical to pioneer communities. Due to the lack of research, however, it difficult to assess the role of individual organizations as pioneer communities. As I discuss in the next section, the rich literature on the history of quantitative practices in journalism illustrates the importance of understanding how older traditions shape newer practices around data, and how this makes data journalists 'pioneers' in respect to these older traditions.

From CAR to data and computational journalism

This section looks at the historical trajectories of quantitative journalism by examining various labels that have been attached to quantitative forms of journalism: computer-assisted reporting (CAR), data journalism, and computational journalism. These labels are often used interchangeably, and the practices they describe overlap. The situation of data journalism is similar to civic tech's, in the sense that there is a "lack of a shared definition of data journalism" (Fink and Anderson 2015, 478) and a general "lack of precision" (Aron Pilhofer quoted in Howard 2014a) when it comes to quantitative journalism.

This lack of precision reflects the state of quantitative journalism as evolving and unstable. After reviewing research on the state of quantitative journalism in various countries across Europe and North America, Fink and Anderson

(2015, 468) suggest that 'computational journalism' (which is sometimes used as a generic term for all kinds of quantitative practices in journalism, see below) is "very much a *field in development* and has not yet solidified into anything resembling a classic Bourdieuean structure with formal poles of cultural, economic, temporal capital". Within newsrooms, the conception of quantitative journalism across countries is "extremely vague, both rhetorically and organizationally" (Fink and Anderson 2015, 479). The different attempts by researchers to delineate and define forms of quantitative journalism tend to be contradictory as well (Loosen, Reimer, and Schmidt 2015, 3).

Despite these difficulties, examining the different terminologies used to describe quantitative practices in journalism is useful for this study for two reasons. First, it helps to understand the various assumptions and norms that underlie those practices, which is important if we are to understand how they might be able to complement the practices of data activists. Second, the different labels listed above reflect historical developments and different understandings for the role of data in journalism. While data journalism appears to be a new trend – a response to more recent developments like open government initiatives, Wikileaks, or discussions about big data in general – an examination of CAR illustrates that the use of data(bases) and discussions about its role in journalism are relatively old. Examining these developments will help us to understand how journalism is responding to datafication. In the following, I discuss how the use of quantitative methods in journalism has meant "different things at different times" (Anderson 2015, 361), and that the "quantitative turn" (Petre 2013) of journalism today has different epistemological roots and leads to different practices.

Computer-assisted reporting (CAR)

What is called CAR today is commonly traced back to what Meyer (2002) coined 'precision journalism' in the 1970s (cf. Coddington 2015; Gray, Bounegru, and Chambers 2012; Parasie and Dagiral 2013). Precision journalism promoted the use of social science methods for collecting and analyzing data in order to perform traditional journalistic tasks more effectively (Coddington 2015; Gray, Bounegru, and Chambers 2012). Meyer's work was significant because it represented an approach to journalism that stood in opposition to the (then famous) 'new journalism' movement, which promoted subjectivity and the use of fictional techniques (Gray, Bounegru, and Chambers 2012). Precision journalism argued for journalism "as a more structural mapping of trends *à la* formal social science", as opposed to journalism as a "narrative telling the story of individuals" (Anderson 2015, 357). By advocating the use of quantitative social science techniques, precision journalism strongly appealed to traditional journalistic notions of objectivity.

Later, Meyer's precision journalism got "recast" (Coddington 2015, 333) as CAR. While largely sticking to the methods and principles described by Meyer (McGregor 2013), computers enabled quantitative analysis at greater scales. This potential to increase the agency of journalists shaped the underlying assumption of CAR: database skills combined with computation would help journalists to cope with the growing complexity and abundance of information (Parasie and Dagiral 2013, 856), and reduce journalism's dependence on press releases or biases towards authoritative sources (Gray, Bounegru, and Chambers 2012). Computerized data analysis and statistics would help to reveal truths 'hidden' in publicly available data, because journalists would otherwise not be able to cope with the volume of information (Parasie and

Dagiral 2013). During the 1980s and 1990s, CAR became closely associated with time-intensive, investigative journalism (Gynnild 2014, 719; Coddington 2015, 334), and has sometimes been used as a label for stories that reveal "injustice in society by pointing out the existence and the causes of a social issue, and identifying solutions to it" (Parasie and Dagiral 2013, 856).

It is important to note that techniques of quantification in journalism have a much longer tradition, which predates precision journalism or CAR (Anderson 2015). Quantitative analysis can be traced back to the 19th century (cf. Bowers 1976). The first use of computers in journalism, according to Cox (2000), occurred in 1952, to predict the outcome of the US presidential election. While they were not the first to introduce quantitative practices to journalism, precision journalism and CAR nevertheless formed the basis for the modern forms of quantitative journalism prominent today. As I describe in the following, while the principles and techniques are similar, both the scope and epistemological underpinnings of these practices have since changed.

Data journalism

The term 'data journalism' has become popular since the late 2000s. Despite the large amount of research on data journalism there is no consensus on a definition (Loosen, Reimer, and Schmidt 2015), and it is disputed as to what exactly differentiates data journalism from CAR. The existing research shows, however, that the appearance of data journalism signals important differences in how data is being used, and how journalists understand its role. This might partly be due to the fact that many 'data journalists' who have entered newsrooms have no connection to the CAR tradition (Parasie and Dagiral 2013, 862). Concerns have been raised that data journalism may be 'divorced' from its history (DeBarros 2010; Howard 2014b).

When data journalism started to become more popular around 2010, researchers and practitioners sometimes described it as a broader application of CAR, or a 'natural evolution' of it that makes use of the sheer volume of data and the new technological affordances for data visualizations and interactive web applications (cf. DeBarros 2010). For example, Gray, Bounegru, and Chambers (2012) state that

> the emergence of the label "data journalism" at the beginning of the century indicates a new phase wherein the sheer volume of data that is freely available online – combined with sophisticated user-centric tools, self-publishing, and crowdsourcing tools – enables more people to work with more data more easily than ever before (Gray, Bounegru, and Chambers 2012, 21).

In a similar vein, it has been argued that CAR is mainly a research method that uses computational techniques to support investigative reporting, while data journalism would make use of modern technologies to integrate data in the "whole workflow of journalism in a fundamental way", with data being the basis for "analysis, visualization and – most important – storytelling" (Lorenz 2010).

However, data journalism is not necessarily just a broader application or extension of CAR. Compared to CAR, at least some data journalists have adapted values and practices that represent an "epistemological break" (Coddington 2015, 335) in several respects. Importantly, traditional CAR journalists and data journalists diverge in their understanding about the role of data in journalism. Parasie and Dagiral (2013) show this for 'programmer-journalists' in Chicago. Instead of trying to reveal 'hidden' stories in publicly available data, these programmer-journalists have a strong belief in 'data transparency' and attempt to disclose truths "through the accessing, combination and processing of complete data" (Parasie and Dagiral 2013,

869). For them, data cannot "lie or hide anything if they are granular, complete and regularly updated" (Parasie and Dagiral 2013, 867). While CAR was leaning towards an understanding of data as "a hidden fragment of information waiting to be uncovered", some data journalists envision data "as a thing that is both massive and already known, where the journalistic value added lies not in the unmasking of a hidden truth but in putting the overwhelming torrent of information into patterned context" (Anderson 2015, 360).

The idea that insights can be gained through access to granular and complete datasets (which is not unique to the programmer-journalists studied by Parasie and Dagiral 2013; cf. Coddington 2015) is compatible with the epistemological principle that underpins datafication, which suggests that insights can now be 'born from the data' (cf. Kitchin 2014, 2). A more important influence, however, is data journalism's connections to open source culture, which are absent in CAR:

> CAR arose out of an effort to marry social science with modern professional journalism, and especially investigative journalism. Data journalism and computational journalism, on the other hand, have arisen from the intersection of professional journalism with open-source culture. (Coddington 2015, 344)

The logic of open participation that is central to open source culture, i.e. of inviting contributions to create collective intelligence (see below), suggests that if granular and complete data is made accessible to readers, they would be able to "extract the meanings of data on their own, and eventually to make their own moral claim" (Parasie and Dagiral 2013, 865).

This implies a very different understanding of the audiences and the publics that data journalists aim to serve. CAR was characterized by a "high-modernist

conception of journalism as a practice that could only be undertaken by those officially sanctioned as journalists" (Flew et al. 2012, 160). It emerged at a time when a more passive vision of the public was typical among professional journalists (Anderson 2011). In this traditional model publics are seen as consumptive, and investigative journalism (to which CAR was closely attached) would primarily rely on the public to "supply the moral outrage that it works to produce" (Coddington 2015, 343; cf. Ettema and Glasser 1998). The 'typical' CAR journalist would thus try to set the agenda for public debate instead of inviting readers to explore the data themselves to draw their own conclusions. By contrast, some data journalists would view readers as "co-constructors of truths and moral claims" (Coddington 2015, 335). Coddington (2015, 343) argues that these differences are illustrated by how impact is measured: the primary measure of impact in CAR was its influence on public institutions or officials, while data journalists would also consider enhancing the publics own understanding as impact.

The relationship between CAR and data journalism shows some interesting similarities to the relationship between RTI activists and open source technologists (see above). CAR is closely linked to investigative journalism and, therefore, to notions of political accountability. What matters most to CAR is not the structure and granularity of data, but its potential to disclose wrongdoings, or structural issues that force authorities to react. Data journalism expanded CAR with its notion of 'data transparency', providing services to users and creating a more collaborative culture. In both data activism and quantitative journalism, the original focus on political accountability was extended by more technical notions of openness and participation, adapted from open source culture.

The distinction between CAR and data journalism, however, remains blurry in practice and the term CAR is still commonly used to describe quantitative practices in journalism (as illustrated, for example, by the importance of the annual conference by the National Institute for Computer Assisted Reporting in the US[7]). The differences between CAR and data journalism are often more gradual and not as pronounced as those among the programmer-journalists studied by Parasie and Dagiral (2013), as numerous studies examining the state of data journalism in different countries illustrate (cf. Karlsen and Stavelin 2014; Appelgren and Nygren 2014; Weinacht and Spiller 2014; Maeyer et al. 2015; Hermida and Young 2017; Fink and Anderson 2015). Still, the shifts in how broadly quantitative techniques are applied, in how the role of data in journalism is understood, what vision of publics guide journalists' work, and the connections to open source culture reflect ways in which journalism is adapting to the new digital media environment that increasingly relies on data.

Computational journalism

Different from CAR and data journalism, computational journalism is primarily understood as applying a way of thinking to journalism, namely what Wing (2006) has coined 'computational thinking'. Computational thinking is "the thought processes involved in formulating problems and their solutions so that the solutions are represented in a form that can be effectively carried out by an information-processing agent" (Cuny, Larry and Wing quoted in Wing 2010). Thinking computationally means using "abstraction and decomposition when attacking a large complex task" (Wing 2006, 33), and

[7] https://ire.org/nicar/.

relies on the "logical, algorithmic, scientific, and innovative dimensions of human cognition" (Gynnild 2014, 723).

Implicitly or explicitly, computational journalism has sometimes been used as a generic term that includes both CAR and data journalism (cf. Anderson 2013b; Loosen, Reimer, and Schmidt 2015; Gynnild 2014; Fink and Anderson 2015). This is unsurprising, as it is implicit in one of the most common definition of computational journalism by Hamilton and Turner (2009). For them, computational journalism is "the combination of algorithms, data, and knowledge from the social sciences to supplement the accountability function of journalism" (Hamilton and Turner 2009, 2). While there are other definitions that attempt to more clearly set computational journalism apart from CAR and data journalism – like the one provided by Diakopoulos (2011), which insists that computational journalism is "distinctive in its focus on the processing capabilities (e.g. aggregating, relating, correlating, abstracting) of the computer" – the fundamental issue with these distinctions is that both CAR and data journalism without any form of computational thinking seem very limited. Indeed, working with structured data arguably always requires at least some form of computational thinking. That Meyer has been described as a pioneer of computational journalism (cf. Gynnild 2014, 723) is, therefore, unsurprising.

Consequently, we can argue that all forms of CAR and data journalism are also forms of computational journalism. However, there are forms of computational journalism which do not necessarily require quantification, and thus do not directly overlap with data journalism or CAR. Examples of these journalistic forms are the automated production of news articles (Carlson 2015), forms of 'algorithmic accountability' reporting (like the analysis of algorithms used in predictive policing by ProPublica, Larson et al. 2016; cf.

Diakopoulos 2015), or tools like DocumentCloud that help to automate common tasks for journalists (Coddington 2015, 336). A counter-argument to using computational journalism as a generic term is the danger of overlooking the distinctiveness of these cases, or missing their overlaps and similarities.

While CAR denotes supporting investigative journalism with computational methods, and data journalism shifts the ways in which data is understood and integrated in professional journalism, computational journalism denotes a more generic way of thinking that is not unique to journalism itself. It is connected to the influence of open source culture on professional journalism, and shows that this influence is not limited to quantitative practices. Computational journalism is thus relevant for this study because the extent to which journalists are capable of thinking computationally, and their willingness to innovate, potentially allows them to extend the "journalistic frames for work" (Gynnild 2014, 723). The ability, or inability, to think computationally also affects how journalists and data activists complement each other and cooperate. In the following, I look into the existing research on this cooperation.

Relationship between data activism and data journalism

As mentioned in Chapter 1, there is a lack of research that is primarily concerned with the relationship between data activists and data journalists thus far. There is research on related issues, however, which can be put into three categories: 1) research on the interactions between journalists and actors from the larger technology sector rooted in open source culture (e.g. professional programmers and computer scientists; in the following abbreviated as

'technologists'); 2) how open source culture is compatible or clashes with professional journalism; 3) the cultural significance of open source culture, and how it shapes fields of knowledge production other than software development. The first two categories consider the influence of technologists on the news-making-process (Domingo, Masip, and Costera Meijer 2015), and the third category focuses generally on the potential of open source culture.

Interactions between journalists and technologists

In Chapter 1, I argue that looking *exclusively* at the direct interactions between technologists and journalists is too narrow if we are to understand how these groups complement each other. Still, looking at those interactions is relevant, because informal events where journalists and technologists meet have had, and continue to have, an important role in bringing these actors closer together. The existing research illustrates under which conditions journalists and technologists (including data activists) are able to collaborate successfully, which might have implications for how they can complement each other more broadly. Moreover, this research describes the broader context in which data journalists and data activists interact.

Particularly important here are two studies by Lewis and Usher (2014, 2016), where they use concepts from science and technology studies to explore the interactions between journalists and technologists at informal events and workshops. The first study describes Hacks/Hackers events as 'trading zones' that offer opportunities to develop mutual understandings (Lewis and Usher 2014). Hacks/Hackers is a transnational grassroots organization that brings together journalists ('hacks') and technologists ('hackers') through informal meetings, to exchange ideas and to potentially collaborate. Lewis and Usher (2014, 391) show how bringing these groups together "requires significant,

coordinated, and sustained effort, and that the barriers between each field's understanding of the other are real". They mention three factors for the success of Hacks/Hackers events: 1) supporting institutions that provide free space to meet; 2) dedicated volunteers that organize regular events; and 3) sufficient cross-understanding (Lewis and Usher 2014). The authors describe that cross-understanding is not only hindered by field-specific jargon, but by "differences in thinking" (Lewis and Usher 2014, 389). For example, journalists tend to hope for solutions to mundane, short-term work, while technologists are interested in long-term software development and investigative scoops.

In their second case study, Lewis and Usher (2016) examine the cross-understandings between journalists and technologists more in-depth. They study an online 'Learning Lab' organized by the Knight-Mozilla News Technology partnership, where participants with diverse backgrounds had collaborated to innovate journalism through open source software. The authors show how journalism's expanding boundaries might lead to new definitions of news. The distinct understandings of news and technology among participants converged around three thematic themes. First, news should be more 'process-oriented'; rather than being limited to a fixed article presented to readers, the news process should "more fluidly and continuously [allow] users to critique, advise, edit, and add upon the work of journalists" (Lewis and Usher 2016, 555). Second, journalism should be more open to audience participation beyond allowing them to comment on news, to the point where "the quality of news is correlated with the range of voices contributing to it" (Lewis and Usher 2016, 554; cf. Lewis 2012b). Third, participants thought that social curation should become more important:

journalism should get better at helping users to identify relevant information (Lewis and Usher 2016, 555).

Lewis and Usher's (2014, 2016) work does not directly contend with activists in the open data or civic tech scene, and it primarily analyzes examples from the US. A lack of studies in Europe about the interaction between technologists and journalists makes it difficult to assess how generalizable these findings are. Still, Lewis and Usher's (2014, 2016) work provides a useful starting point, as many of the technologists drawn to journalism are coming from the open data and civic tech movement (Parasie and Dagiral 2013), and the meaning of many key terms around data journalism and civic tech has strongly been influenced by developments in the US (see above).

Ananny and Crawford's (2015) study of news app designers (e.g. developers working at Google News) looks beyond direct interactions and explores the professional identities of actors at the intersection of the sectors of journalism and technology. They suggest that news app designers constitute an emerging 'liminal press': "a set of field-level relationships among actors who may not self-identify as journalists but nonetheless define the conditions under which news is created and circulates" (Ananny and Crawford 2015, 193). News app designers work in a space between the fields of technology design and journalism, and are "influenced by both but not entirely beholden to either" (Ananny and Crawford 2015, 204). Emphasizing that "the understanding of news shifts depending on the industry that is delivering it" (Ananny and Crawford 2015, 196), the authors found that the artifacts produced by news app designers serve as boundary objects where app designers attempt to mark their work as different and distinct from the work of journalism. Ananny and Crawford (2015, 204) suggest that the 'boundary work' between journalism and the technology sector does not only coalesce around particular objects or

events where journalists and technologists directly interact, but also along a whole 'boundary infrastructure' across multiple actors, objects and practices: "These actors build and sustain boundary infrastructures because they can be read as members of an identifiable field (mobile, personalized news app designers) – but they also work in sufficiently diverse ways to appear different from their competitors and traditional, mainstream journalists".

While we cannot easily compare the practices, self-understandings, and identities of data activists working in non-profits with those of news app designers working in for-profit companies, Ananny and Crawford's (2015) study does illustrate the need to look beyond the direct interactions between data journalists and data activists, and take a broad range of boundary-spanning objects into account. In addition to this need, their study shows that the self-understanding and positioning of these actors is key for assessing the entanglements between them. As I discuss in the next section, open source culture plays a particularly important role for the identities of data journalists and data activists, and how they position themselves vis-à-vis each other.

Open Source culture and professional journalism

As mentioned above, modern forms of quantitative journalism "have arisen from the intersection of professional journalism with open-source culture" (Coddington 2015, 344). Lewis and Usher (2013, 604) argue that open source culture was crucial to the "project of bringing together journalists and technologists", as many technologists who entered newsrooms with the adoption of data and computational journalism adhere to open source principles. Research about the resulting renegotiation of values and practices between technologists and journalists has concentrated on the implications of open source culture for professional journalism. This work asks how open

source culture is integrated into newsrooms, and explores the implications for the boundaries and the identity of journalism.

Research about the relationship between open source and journalism shows that the values and practices of open source culture do not always easily align with the practices, identities and values of professional journalism. As Lewis (2012b) describes, institutionalized journalism is shaped by a professional logic of (exclusive) control over content, while open source technologists adhere to a logic of open participation that relies on the 'wisdom of the crowds'. On the one hand, this suggests that an introduction of open source culture in journalism might result in a questioning of the "ideal of journalism as a profession" (Josephi 2012, 481). On the other hand, we would oversimplify the complex relationship between open source and journalism by stating that an adoption of open source culture by journalists weakens their professional autonomy. Research into data journalism also indicates quite the opposite: many accounts of data journalism suggest that gaining expertise in working with data is used to "justify an authoritative journalistic role in which the journalist can answer the public's questions about data" (Lesage and Hackett 2014, 41), which might result in a strengthening of professional autonomy.

Professional journalism regularly falls into a pattern of 'normalizing' open source culture to fit it into traditional practices, identities and routines. The concept of normalization describes journalisms' tendency to 'absorb' potentially disruptive practices in ways that secure the professional autonomy of journalism as the gatekeeper of publicly relevant information (Singer 2005). Studies of data journalism have shown a clear pattern of subordinating technical skills "to the unbroken tradition of journalism" (Karlsen and Stavelin 2014; cf. Royal 2010). Technologists tend to embrace journalism (Royal 2010), and in their work they largely focus on developing tools that help traditional

journalists to "further their goals of doing journalism the way it has always been done" (Lewis and Usher 2013, 609). Values and practices of open source, such as collaboration, sharing, and transparency, tend to be adapted in ways that preserve journalism's gatekeeper role. For example, rather than inviting readers to more actively take part in the production process of news, journalists have developed tools that facilitate the collaboration and exchange among fellow journalists (Baack 2016).

Lewis and Usher (2013, 609) criticize that journalists and technologists "have largely failed to interrogate the old processes of newswork". They propose a more radical application of open source culture to journalism, which would make the news-making process more participatory, collaborative, and transparent. The application involves imagining news stories as code, and reimagining journalism as knowledge management, a proposal quite similar to Hansen's (2012) call for a new 'aporias' in journalism. Both Lewis and Usher (2013) and Hansen (2012) suggest that journalism should be more process-oriented. According to Hansen (2012), in the digital age, journalism should be understood as a critical form of moderation that both gives and denies room for dialog. The journalist should primarily act as the "editor, moderator and curator of the information flow" (Hansen 2012, 680). Lewis and Usher's (2016) later empirical work shows that similar re-articulations and understandings are indeed emerging at the intersection between technologists and journalists (see above).

Yet, while Hansen (2012) and Lewis and Usher (2013) constructively illustrate the potential influence of open source culture on journalism, the empirical question remains as to whether or not, and how, journalism is structurally changing with the influence of open source culture. Moreover, Hansen's (2012) and Lewis and Usher's (2013) work maintains a focus on changes within

newsrooms, and does not consider how data journalism affects data activists, or how the practices and values of data activists complement professional journalism *without* becoming part of newsroom culture. While open source culture is certainly important for data activists, we cannot simply equate their practices and values with open source, as I discuss in the next section.

Open source culture and data activists

Research on the relationship between journalists and technologists, implicitly or explicitly, assumes that the practices and values of technologists can be equated with open source culture. As researchers who are interested in the cultural significance of open source have shown, however, open source culture is continuously re-articulated or "modulated" (Kelty 2008) when it is applied to domains outside of software development. This raises the question of how open source culture not only influences journalism, but also data activists themselves, and what the ways in which they modulate open source might mean for the relationship between data activists and journalists.

Open source culture is rooted in hacker culture, the latter of which adheres to a set of values and beliefs that emerged in informal networks of cooperation for the free and creative use of technology (Levy 1984). The technical and ethical commitments of early communities formed an important reference point from which hacker culture then diversified into a broad range "of distinct yet connected moral genres" that draw from liberalism (Coleman 2013, 19). Open source culture "derived from a more positive notion of liberty" (Coleman and Golub 2008, 261), around collaboration and sharing. As mentioned above, the key principle of open source culture is voluntary participation (Weber 2004, 62): not only is the source code openly available, but potentially everybody is invited to contribute to it. Since its origins in the

1980s, open source culture has spread and diversified into a great variety of contexts. For example, open access advocates for more transparency and collaboration in academic publishing; and Creative Commons has developed copyright licenses inspired by open source for all kinds of content.

In his study about the cultural significance of open source culture, Kelty (2008, 57) suggests that open source is about "operating systems and social systems: the imagination of order shared by geeks is both moral and technical". Open source is a combination of architecture and culture (Lewis and Usher 2013, 606) in the sense that it is comprised of practices for building infrastructures that embody the ethos of sharing and collaboration. As a result, open source advocates create 'recursive publics', in which "geeks use technology as a kind of argument, for a specific kind of order: they argue about technology, but they also argue *through* it. They express ideas, but they also express *infrastructures* through which ideas can be expressed" (Kelty 2008, 29).

Put differently, recursive publics created by open source advocates differ from 'normal' publics because they constitute themselves not only through the circulation of (objects of) discourse, but also through the creation, modification, and maintenance of the infrastructures through which discourse can circulate. Recursive publics are collectives "*independent of other forms of constituted power and (…) capable of speaking to existing forms of power through the production of actually existing alternatives*" (Kelty 2008, 3). While this independence is not absolute, it is a key factor that helps to explain the political power of hacking in general, as Coleman (2016, 166) points out. She also notes that recursiveness, in this sense, has a counterpart in what sociologists of social movements call 'free spaces': settings which are removed from direct control of dominant groups, and therefore "generate the cultural challenge that precedes or accompanies political mobilization" (Polletta 1999, 1).

According to Kelty (2008), projects and groups that derive from open source culture, such as Creative Commons (Garcelon 2009) or Wikipedia (Tkacz 2015), apply the key practices and ethos of open source to new domains. Because open source practices have to be 'modulated' in these processes – for example when Creative Commons applies open source inspired copyrights to 'content' rather than source code – Kelty (2008) calls such practices 'modulations' of open source. These modulations are "committed to experimenting with the given practices of Free Software" (Kelty 2008, 246), to alter the existing systems of knowledge production.

Following Kelty's (2008) classification, open data and civic tech are modulations of open source culture. Due to a lack of research on the practices and self-understandings of data activists, however, it remains unclear how civic tech and open data modulate open source culture, what role data plays in these modulations, and what the implications are for how activists imagine the publics they aim to address. Moreover, recursiveness raises interesting questions about the relationship between data activists and data journalists. On the one hand, recursiveness implies that the extent to which journalists can normalize the practices and values of data activists is limited, because activists can create 'actually existing alternatives' that challenge the professional boundaries of journalists 'outside' of their own profession. On the other hand, activists' ability to form recursive publics might also decrease the pressure on journalists to weaken their boundaries, as activists can complement or extend the work of journalists independently without necessarily challenging the traditional routines inside of newsrooms. These questions providing directions for the empirical analysis in Chapter 4-6.

Implications: Datafication and monitorial democracy

The ways in which data activists and some data journalists understand their work, and the public services that they aim to provide, resonate with how developments in democratic practice and discourse have been theorized by Schudson (1998, 2015) and Keane (2009): an evolution from informed to 'monitorial' forms of citizenship and a related change from representative to 'monitorial democracy'. By illustrating the links between monitorial democracy and the history of quantitative practices in journalism and around civic tech, I aim to show that data journalism and data activism can be considered related phenomena with common roots and similar implications. This analysis further underlines the argument made in Chapter 1: that commercial actors and governments have not been the only drivers of datafication. First, I describe monitorial citizenship and monitorial democracy, and how they resonate with data activism and data journalism. Then I review how these concepts have been applied to data journalism and data activism by others, and consider the implications for this study.

Monitorial democracy and datafication

Schudson (1998, 310) first developed the concept of 'monitorial citizenship' in his study about the history of citizenship in the US, where he concludes that the classic ideal of the 'informed citizen' is inadequate for the full understanding of citizenship in modern democracies: "if democracy requires omnicompetence and omniscience from its citizens, it is a lost cause. There must be some distribution across people and across issues of the cognitive demands of self-government". We should, rather, understand modern forms

of citizenship as 'monitorial', because the average citizen "scans (rather than reads) the informational environment" (Schudson 1998, 310) and only raises her voice if necessary. Monitorial citizens tend to be "defensive rather than pro-active" (Schudson 1998, 311), and rely on informed 'watchdogs' to become aware of relevant issues that require their intervention. Schudson (1998) does not suggest that this model of citizenship is superior or inferior to other models, but he rejects "idealized textbook notions of the democratic public sphere in favor of close attention to the messy and invariably disappointing realities of journalism and political discourse" (Graves 2017).

Keane (2009; also discussed by Schudson 2015) expanded upon these ideas, arguing that representational forms of democracy have been slowly turning into 'monitorial democracies' since 1945. Keane (2009) suggests that a systemic distrust in representational practices has led to an intensified monitoring of power. Representative mechanisms are increasingly "mixed and combined with new ways of publicly monitoring and controlling the exercise of power" (Keane 2009, par 1.80). For Keane (2009), this is evidenced by an ever-increasing number of power-monitoring devices or groups, e.g. democratic audits, human rights organizations, consumer councils, or online petitions. These "new mechanisms of representation (…) allow continuous, rather than episodic, representation; popularly generated rather than party-controlled representation; and many platforms for entrepreneurial democratic action" (Schudson 2015, 241).

Similar arguments have been made by Rosanvallon (2008), who suggests that distrust is the defining characteristic of modern democracy. What he calls 'counter-democracy' is "a form of democracy that reinforces the usual electoral democracy as a kind of buttress, a democracy of indirect powers disseminated throughout society (…) a durable democracy of distrust, which

complements the episodic democracy of the usual electoral-representative system" (Rosanvallon 2008, 8).

Regardless of whether we call it monitorial democracy or counter democracy, these authors describe a similar phenomenon: that a distrust in conventional practices of representation has sparked a multiplicity of monitoring actors that create a mode of governing, "in which power is everywhere subject to checks and balances, such that nobody is entitled to rule without the consent of the governed, or their representatives" (Keane 2009, par 9.133).

When we compare these ideas with the history of data activism and data journalism described above, we can see several links. Similar to monitorial democracy, RTI and the application of open source culture to institutional politics make a case for the continuous monitoring of representational mechanisms. Ultimately, RTI and open source similarly aim to ensure that a multiplicity of voices are present in decision-making processes within and beyond institutionalized practices of representation (e.g., periodic voting). The difference is in the means: RTI suggests that the rights of citizens to acquire information is of primary importance; open data and related expressions of open source culture, alternatively, suggest that governments should be more transparent and collaborative. The ways in which practitioners attempt to define civic tech as "any tool or process that people as individuals or groups may use to affect the public arena" (Sifry 2014), or as "improving methods of getting public input into government" (Shaw 2016), are very compatible with Schudson's (1998) notion of monitorial citizens. A similar pattern has emerged in quantitative journalism. Monitoring governments is at the heart of investigative journalism, with which CAR is closely associated. The way some data journalists think of their audience as "co-constructors of truths and moral claims" (Coddington 2015, 335) expands upon the practices and

epistemologies underlying the CAR tradition with practices and values from source culture to promote transparency and collaboration. While they differ in their emphasis on professionalism and collaboration, both CAR and data journalism can be said to support monitorial forms of citizenship.

The connections between data activism and data journalism to Schudson's (1998) and Keane's (2009) work point to a more fundamental relationship between quantification and monitorial democracy. Quantification and numbers are an important part of monitoring government performance, and in ensuring that forms of participation are legitimate in the sense that they are representative (Rose 1999, 200–201). The premise of monitorial democracy is a growing demand for 1) monitoring and scrutinizing governments that claim to be democratic, and 2) more continuous forms of participation beyond institutionalized practices of representation. At the very least, these premises suggest that practices of quantification become more widespread and more central to democratic discourse. Whether or not we agree with the premises of monitorial democracy, data journalism and data activism can be understood as different responses to perceived changes in the demands of the public, as both of them arguably support monitorial forms of citizenship in various ways. In the following, I show that this has also been discussed by other researchers.

Monitorial democracy applied to data journalism and data activism

Given that both data activism and data journalism are compatible with the premises of monitorial democracy, it is unsurprising that the concept of monitorial citizenship has already been applied to data activism and the changing role of journalism. Civic tech has been explicitly theorized in terms of monitorial citizenship by Zuckerman (2014a, 2014b; Hope 2014). He

suggests that the applications of civic technologists – e.g. parliamentary monitoring software, problem reporting, or freedom of information websites – are tools that help monitorial citizens to be more efficient and effective. For Zuckerman (2014b, 156), civic tech is related to other potentially monitorial practices, such as online petitions or crowdfunding, which he understands as 'participatory civics': forms of civic engagement and activism that "embrace a post-'informed citizen' model of civic participation". According to his model, forms of participation can seamlessly move from rather 'thin' modes of engagement (such as expressing support in an online petition) to 'thick' modes of engagement (such as marching in a demonstration). Civic tech would often make thin modes of engagement easier, but also helps to connect 'thinner' actions with 'thicker' ones, e.g. when participating in an online petition leads participants to directly contact their representatives. Zuckerman's (2014a) work is one of the most prominent ways to theorize civic tech to date, and has been influential within the civic tech sector. Schrock (2016, 592) has also suggested that civic technologists themselves form a 'monitorial elite' that is "watching data streams and processes of algorithmic regulation for injustices and engaging directly with local politics".

While monitorial citizenship has not been applied to quantitative journalism specifically, it has also been used in debates within journalism studies about the changing role of the profession. For example, Deuze (2008) suggests that monitorial citizenship requires a new paradigm for journalism in general. For him, monitorial citizenship signals a more individualized form of citizenship, in which citizens act like consumers who participate in public issues "conditionally, unpredictably, and voluntarily" (Deuze 2008, 852). Therefore, journalists could no longer assume that the old model of journalism is still adequate (if it ever was), in which they provided news to 'the masses' in order

to transform them into informed citizens. Instead, journalism should act as "an amplifier of the conversation society has with itself" (Deuze 2008, 848). The similarities to Lewis and Usher's (2013, 2016, 2014) and Hansen's (2012) work described above are striking. Both in their theoretical approaches and in the empirical evidence that these authors present, the adoption of data and computational techniques and the influx of technologists into newsrooms point to a re-articulation of journalism to be more process-oriented and participatory.

As monitorial democracy describes a shift towards "a variegated institutional landscape in which the lines between politicians, journalists, experts, and everyday citizens are not always clear" (Graves 2017, 1244), some authors have also considered the changing relationship between journalism and civil society through the analytical lens that this concept provides. As an example, Breindl (2016) suggests that activists increasingly take over monitorial watchdog roles, which have traditionally been attributed to journalists. Because they often specialize in monitoring particular issues, activists would increasingly emerge "as *experts* and *legitimacy sources* for decision-makers and journalists, who trade access to the policy and media system in exchange for information and public support" (Breindl 2016).

While Breindl (2016) makes some interesting observations, her study is not easily applicable to the entanglements between data activists and data journalists. Data activists do not necessarily act as 'experts' for specific subjects; they aspire to change the relationship between governments and their publics more broadly. A consideration of data journalism and data activism as related phenomena is still missing from the current research literature.

Conclusions

As this thesis asks, first, what role data plays in the social imaginaries and practices of data activists and data journalists, and second, how their practices and imaginaries diverge and converge, the previous sections have several implications for the methodological design and empirical analysis:

1. We need to be sensitive to the nuances and differences within quantitative journalism and data activism, and how they affect entanglements between these groups

The historical trajectories illustrate that the imaginaries and backgrounds of data journalists and data activists are diverse. This might affect their ability to complement each other, or directly collaborate. For example, it might be easier for journalists closer to the CAR tradition to develop cross-understandings with activists from an RTI background, as both parties have a stronger emphasis on political accountability and the specific quality of information. Rather than treating them as coherent and fixed entities, we should ask whether we can identify different groups within these fields that use and understand data differently, and the ways in which this shapes how they are entangled.

2. We need to understand how data activists rely on open source culture, and how this shapes their use of data and their ability to collaborate or complement data journalism

As shown above, discussions about the relationship between journalists and technologists tend to implicitly or explicitly equate technologists with open source culture, even though research on the spread and diversification of open source culture demonstrates that we cannot simply rely on notions of 'hacker culture' or open source software development to understand the role of such actors. When applied to domains outside of software development, open

source culture becomes re-articulated and 'modulated' (Kelty 2008). For journalism, the literature shows that open source culture is an 'externality' that tends to get normalized, i.e. adapted in ways that preserve journalists' role as gatekeepers for publicly relevant information. Data activism, however, in part grew directly out of open source communities. This raises questions about data activists that have not been addressed thus far. How important is open source culture to the practices and imaginaries of data activists? How do data activists modulate open source, and how does this shape their use of data and their ability to cooperate and complement journalism?

3. We should grasp data journalism and data activism as related phenomena, and consider the implications of their shared roots and entanglements

The literature review confirms the critique made in Chapter 1: while it is largely acknowledged that data activists and data journalists are entangled and have overlaps, researchers continue to theorize them as clearly separated phenomena. They look at how data journalist and data activists interact in joint workshops and projects, or how actors at the intersection of journalism and the technology sector understand their work (what Ananny and Crawford 2015 call a "liminal press"). By showing how monitorial democracy is linked to both data activism and data journalism, this chapter further underlines the argument that we have to consider the implications of their entanglements, rather than how each 'field' is changing due to its increased reliance on data. For the empirical analysis of this thesis, the necessity to look across institutional and organizational boundaries also implies asking how applicable the concept of monitorial democracy is to data journalism and data activism, and how it does or does not help to examine and theorize the implications of the entanglements of data activists and data journalists.

In the next section, I describe the methodological approach of this study. This approach takes the insights gathered in this chapter into account, and aims to be sensitive to the nuances within and across data journalism and data activism.

3. Methodology

Studying pioneer communities through practices

This chapter outlines and justifies the methodological approach of this study. First, I explain how the focus on practices used to study data journalists and data activists as pioneer communities was implemented methodologically following a constructivist grounded theory approach (Charmaz 2006). Second, I describe the different methods used to collect data: mapping the field through participant observations and Digital Methods, interviews, participatory mapping, and content analysis. Third, I describe how I applied constructivist grounded theory in the three case studies. I conclude with an evaluation of my methodological approach.

A constructivist grounded theory approach

To study pioneer communities through practices, the methods used in this study are mostly qualitative because they enable me to be open-minded about "what people are doing and how they categorize what they are doing" (Couldry 2004, 125). Because studying the social imaginaries of data activists and data journalists requires an understanding of how they make sense of their own practices, I followed constructivist approaches in the tradition of cultural studies or ethnomethodology. According to Alasuutari (1996), these approaches do not aim at achieving generalized understandings of universal mechanisms that would help to explain what people 'really' mean or think. Instead, cultural studies or ethnomethodology are interested in revealing a "local and historically specific cultural or 'bounded' system" (Alasuutari 1996, 371) by studying how "participants interpret each other's speech and thus

achieve a shared understanding of the situation" (Alasuutari 1996, 382). In this sense, the goal of theories inspired by cultural studies or ethnomethodology is to deconstruct the ways in which "we construct realities and social conditions and ourselves as subjects in those realities" (Alasuutari 1996, 382).

To guide the data collection and analysis, I rely on a variant of grounded theory which explicitly aims at developing cultural theories in the sense outlined by Alasuutari (1996): constructivist grounded theory (Charmaz 2006). Grounded theory is both a set of flexible guidelines for collecting and analyzing qualitative data, and the outcome of a research process that follows these guidelines (cf. Charmaz 2006, 3–4). It was chosen for this study because of its openness as an ongoing learning process based on the principle of theoretical sampling. Theoretical sampling means that an initial data sample is continuously expanded to further elaborate and refine the evolving theoretical concepts, causing data collection and analysis to overlap. Through theoretical sampling, the researcher cycles through several 'spirals'. Each passage through the spiral of grounded theory includes the same basic steps of collecting data, coding, and memo writing, but with each theoretical sampling the process starts from a 'higher' (i.e. more developed) theoretical vantage point (cf. Krotz 2005, 167). This cyclical progression is meant to "tighten (…) the corkscrew or the hermeneutic spiral so that you end up with a theory that perfectly matches your data" (Hood quoted in Charmaz 2006, 101).

This makes grounded theory very compatible with practice-focused research paradigms that strive to look beyond pre-defined categories, because it facilitates the discovery and development of ideas and concepts that may have initially been less obvious, or even have appeared to be counter-intuitive. Grounded theory also helps to develop nuanced understandings of people's

own interpretations, as it provides techniques designed to question preconceived ideas in order to develop theoretical sensibility (see below).

Following the introduction of the basic techniques and concepts of grounded theory by Glaser and Strauss (1967), several variants of grounded theory have emerged that differ in both technique and epistemological underpinnings (for an overview, see Hallberg 2006). In this study, I rely on *constructivist* grounded theory, a variant developed by Charmaz (2006). Charmaz (2006) aims to change the epistemological foundations while preserving some of the methodological concepts from earlier variants of grounded theory. For her, basic grounded theory guidelines such as coding, memo-writing, constant comparison, and theoretical sampling are neutral, but the ways in which researchers use these guidelines "is not neutral; nor are the assumptions they bring to their research and enact during the process" (Charmaz 2006, 9).

While constructivist grounded theory adopts many of the techniques introduced by Glaser and Strauss (1967) and later by Strauss and Corbin (1998), Charmaz (2006) criticizes some earlier grounded theorists for their positivism. They understood grounded theory as a "method of discovery, treated categories as emergent from the data, relied on direct and, often, narrow empiricism" (Charmaz 2006, 8). Charmaz (2006) calls the application of grounded theory methods following this epistemological paradigm 'objectivist grounded theory'. Objectivist grounded theory assumes that data represent "objective facts about a knowable world" (Charmaz 2006, 131). The researcher's role "becomes more of a conduit for the research process rather than a creator of it" (Charmaz 2006, 132).

By contrast, constructivist grounded theory does not assume that theories 'emerge' from data and describe an objective reality. It "sees both data and

analysis as created from shared experiences and relationships with participants" (Charmaz 2006, 130). Neither the researcher's insights, nor the perspectives of research participants, are considered to be objective facts. Similar to the distinction made by Alasuutari (1996), objectivist grounded theories are implying to reveal universal truths (or, in the language of classic grounded theory: 'basic social processes'), while constructivist grounded theories aim to "provide interpretive frames from which to view realities" (Charmaz 2006, 128). Charmaz (2006, 184) describes constructivist grounded theory as a return to the pragmatist foundations of grounded theory, that aims to construct an "*interpretive rendering* of the worlds we study rather than an external reporting of events and statements". Constructivist grounded theory puts the emphasis on "*how*–and sometimes *why*–participants construct meanings and actions in specific situations" (Charmaz 2006, 130), and therefore leans towards theorizing actions and fluid processes rather than static concepts.

This sensitivity for how meaning is constructed makes constructivist grounded theory very suitable for studying the ways in which data journalists and data activists interpret their practices, and how they relate to each other. Developing 'interpretive renderings' of their modes of understandings is important to this study for two reasons. First, it is necessary to identify overlaps and contradictions in the social imaginaries of data journalists and data activists, and the ways in which these overlaps and contradictions might shape the entanglements between them. Second, their modes of understanding and the role of data in those understandings is relevant given their status as pioneer communities: their own understandings and practices constitute the 'horizon of possibility' that provides orientation for others (see Chapter 1). In

the following, I first discuss the methods I used and then describe how I applied constructivist grounded theory.

Methods

As mentioned above, this study employs a primarily qualitative, multi-methodological approach to examine the practices and social imaginaries of data activists and data journalists, and the ways in which both groups overlap and interrelate. In this section I describe the following methods: semi-structured interviews, content collection and analysis, ethnographic work, and Digital Methods.

Following grounded theory's logic of theoretical sampling, the methods discussed here complement each other, and were not conducted in isolation: ethnographic work and content found online informed the selection of interviewees and the interview guides for the initial sampling; the interviews also helped to select sites for ethnographies and relevant content.

Mapping and accessing the field: Ethnographies and Digital Methods

Mapping the field through ethnographic work and Digital Methods (Rogers 2013) plays an important role in this study for several reasons, even though most of the insights derived from the data that was generated by those methods is not directly referred to in the three case studies presented in Chapters 4-6. First, ethnographic work and Digital Methods informed the selection of content and interviewees. Second, they contextualized and enriched the interviews, as I was able to gain a deeper understanding of the techniques used by my interviewees, and receive an impression of the larger

international communities that they are part of. Third, the use of ethnographic work and Digital Methods has played a role in the evaluation of my methodological approach (see evaluation below).

Ethnographic miniatures

Participant observations are useful extensions of interviews and content analysis because they enable the researcher to see how groups interact with each other, how they present themselves, and how they work (individually or collaboratively). Moreover, participant observations allowed me to gain practical knowledge myself, for example by learning how to write my own scrapers, clean data, and do some analysis using the same tools as my interviewees. This practical knowledge proved to be vital to the understanding of the practices of data journalists and data activists, and it made the interviews more effective; as it enabled me to not only follow my interviewees, but also ask them more specific questions about the tools that they used or issues that they faced. The ethnographic participation included popular online teaching courses for data journalists or data activists, most importantly *Doing Journalism with Data: First Steps, Skills and Tools*, which was organized by the European Journalism Centre.[8] I also took several introductory courses to programming that were not directly related to data journalism or data activism, for example *Learn Python the Hard Way* and *An Introduction to Interactive Programming in Python*.[9]

[8] http://learno.net/courses/doing-journalism-with-data-first-steps-skills-and-tools.

[9] See https://learnpythonthehardway.org/, https://www.coursera.org/learn/interactive-python-1 and https://www.coursera.org/learn/interactive-python-2.

In addition to taking online courses, I visited the following four conferences (in chronological order): The Open Knowledge Festival 2014 in Berlin, the Daten Labor conference 2015 in Dortmund, the Mozilla Festival 2015 in London, and the netzwerk recherche Jahrestagung 2016 in Hamburg. I also participated in several workshops and meetups in Berlin on a regular basis in 2016 and 2017, e.g. Hacks/Hackers Berlin, #DDJ Berlin, Data Visualization meetup Berlin, Code for Germany Berlin. In classic ethnographic fashion, my role during these conferences and meetups was what Brennen (2013, 165) refers to as "participant as observer". I tried to integrate in the activities and discussions at those events as thoroughly as possible, while making field notes about observations and experiences. Collectively, these participations form what Backmann and Wittel (2011, 191) describe as 'accumulated ethnographic miniatures'. Rather than staying in one field for a longer period of time, which is the classic notion of ethnography, I combined several short stays in different places.

Most of the data collected during these stays expanded and informed the interviews and the selection of content. The goal of the ethnographic miniatures was to explore and map the terminologies, practices, concepts, and issues that data activists and data journalists are concerned with, how they interact with each other, and what backgrounds they have. The data included notes about observations, keywords, impressions, and theories. For example, I collected project names, subjects discussed by participants, and the issues that they were working on. Additionally, I took pictures of where and how people interacted. These pictures helped to capture the locations and settings in which the participants met, how they interacted with each other in physical meetings, and how they worked. Participating in those events was also

invaluable in terms of networking within local or national communities (for a more detailed description of the data collection and analysis, see Appendix A).

Digital Methods: Scraping civic tech organizations on GitHub

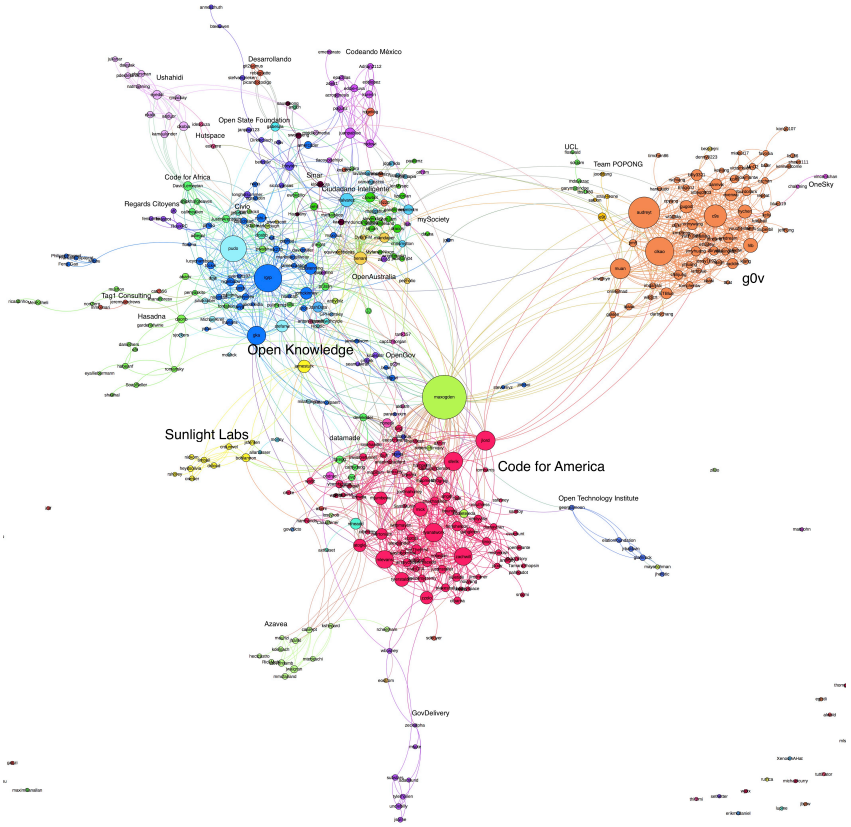


Figure 1: Follower network of civic tech organizations on GitHub.

In addition to qualitative methods, I conducted a quantitative analysis of civic tech organizations on GitHub, an online platform for collaborative software development commonly used by civic technologists. The goal was to gain a deeper understanding of the civic tech community on a global scale, and to better understand the role of the two organizations that I studied in-depth: the OKF DE and mySociety. Gathering and analyzing GitHub data was both inspired by and designed to complement a project that I contributed to during

the Digital Methods Summer School 2015 at the University of Amsterdam[10]. The project was organized by Liliana Bounegru, Jonathan Gray, and Stefania Milan, and explored how GitHub serves as a transparency device for 'journalism in the making', and how it mediates, challenges, or transforms journalism practices, products, and values (Bounegru 2015). The data gathered from GitHub can show who contributes to a software project and for which organizations these individuals work. The scraper also provides data to analyze the use of GitHub's social features, such as following other users or 'watching' software repositories. To gather this data, I developed a scraper written in Python that collects data via GitHub's API.[11]

The scraper produces spreadsheets showing, among other things, how many 'watches' software repositories of civic tech organizations have, the locations of civic technologists, as well as network graphs that illustrate follower and contributor networks across organizations (which can be analyzed with the network analysis tool Gephi). I published this data along with some preliminary analysis on my personal homepage and posted it in the official discussion forum of mySociety.[12]

The project received a lot of attention in the broader civic tech community, and I received suggestions for further analysis. Some civic technologists also shared their interpretation of their own position in the networks, which added

[10] https://wiki.digitalmethods.net/Dmi/SummerSchool2015.

[11] The scraper can be found at https://github.com/sbaack/github-scraper.

[12] https://groups.google.com/a/mysociety.org/forum/#!msg/mysociety-community/DyCFE7bk4_U/Diq9f9kqDwAJ.

qualitative richness to the data. In addition, the administrators of the forum of Poplus, a global community of civic tech organizations, provided me with an internally crowdsourced list of members which included many civic tech organizations that I was not formerly aware of. This information catalyzed a follow-up post, in which I re-ran the scraper with the more complete list of civic tech organizations (Baack 2015b). The attention this follow-up blog post received continued to be helpful for the data collection and analysis of this thesis, as more people became familiarized with my work, and interested in sharing their perspectives with me.

In-depths interviews and participatory mapping

At the heart of this study are in-depth, semi-structured interviews with data activists and data journalists (Weiss 1994; Kvale 1996). Interviews are a common tool in qualitative research to acquire "rich, dense, thick descriptions" (Kvale 1996, 233) from those involved in the life worlds which are being studied. In total, 29 interviews with 27 interviewees were conducted, most of which (22) were face-to-face in Germany or in the Netherlands, and the rest conducted via Skype. The interviews lasted between 30 minutes and two hours; the average length was one hour. The interview guides varied for the different case studies, and customized questions were prepared for each interviewee, using information and documents available online. However, I consistently explored some key subjects and applied two techniques: reconstruction interviews and participatory mapping.

Crucial to most interviews was the reconstruction of a particular project.[13] Each interviewee was asked to pick a project that illustrates her or his work and then walk me through the development process: What was the initial idea behind the project, what were the different steps and phases to implement the project, what happened after the initial release? Reich (2013, 422) refers to these as 'reconstruction interviews', because they reconstruct "technological 'biographies'". Exploring the development of a project in-depth vividly illustrated the practices and routines of the interviewees. If not mentioned by the interviewees, specific follow-up questions about the role of data were asked. After we discussed one project in-depth, I asked my interviewees whether this was a typical project and if there are contrasting examples. If there were, I explored those as well.

Another crucial element to most interviews was what Emmel (2008) refers to as participatory mapping. The interviewees were given a blank piece of paper (or a link to a web application) and asked to draw a network with all the communities that they feel they belong to and other groups, websites, or tools that influence their work. It was suggested, though not prescribed, that they plot themselves in the middle and draw the entities around them. Participatory mapping is usually employed within the context of an in-depth interview as a "name-generator" (Edwards 2010, 8), but, importantly, is also used to explore the meaning of the relationships the interviewees have with the different

[13] Reconstructing projects was important in interviews with developers, project managers, and data journalists. For other interviewees, such as the CEOs of civic tech organizations, I skipped reconstructing individual projects and concentrated more on the founding and direction of the organization.

entities they plot in the network. During the drawing process, interviewees were asked how they relate to the different entities they mention, and how they define their own work in relation to them. If the interviewees did not mention journalists or data activists themselves after they finished drawing the map, I explicitly asked where they would locate them and if they play any role for them. Employed in this way, the process of participatory mapping helped to reconstruct their subjective sense of belonging, and processes of communitization across organizational boundaries (cf. Hepp, Berg, and Roitsch 2014); e.g. when a data journalist feels connected to a group of civic technologists in various ways. The participatory mapping approach also assisted in the assessment of how the interviewees position themselves professionally, and understand their own role in the public arena more broadly.

Other themes addressed in each interview were: self-understandings (preferred job title, understanding of civic tech or data journalism), the ways in which values and ambitions influence the design and development of tools (mostly data activists) or the selection and the telling of stories (data journalists), as well as personal and/or organizational ambitions and values. For more details, see an exemplary interview guide with comments in Appendix B.

Content collection to inform interviews and expand data analysis

This research project profited from the openness of the interview subjects, and their efforts to curate their online presence: the organizations in which they work are often very explicit about their mission and history on various websites (main homepage, project specific websites etc.), and almost every interviewee had given interviews elsewhere, delivered presentations at conferences with videos or slides available online, or written about her or his

work on a personal blog or elsewhere. Collectively, this data provided a rich source of information that was used extensively. First, the materials aided with the selection process of interviewees, especially regarding second research question ('How do the practices and imaginaries of these actors diverge and converge, and how does this shape the entanglements between them?') where I conducted interviews with data journalists. The collected material helped to identify 'leading voices' in the data journalism community, what different viewpoints exists within this community, and who should be interviewed to cover these different viewpoints. Second, materials were collected to prepare the interviews, e.g. by collecting and analyzing previous interviews or relevant blog posts. Often, the collected material already covered parts of the interview guides. In these cases, I prepared more specific follow-up questions, and referred to the source material during the interview.

Third, some of the collected material was used in the post-interview data analysis, both in initial and theoretical samplings. For the case studies on the OKF DE and mySociety, several self-portraying documents, such as the official homepage and project websites, were used to start the analysis with initial coding (see Appendix A). For the second case study about mySociety, the Wayback Machine of archive.org was used to retrieve earlier versions of the homepage, which has been online since 2003. In addition to providing important insights into the organization's self-understanding, retrieving old announcements or older versions of project-specific websites also served to complement the reconstruction interviews. After the initial sampling, additional documents were included in the analysis, following theoretical samplings. In total, 38 documents were included in the final analysis (these included self-portraying websites, interviews given by interviewees elsewhere,

transcripts of recorded presentations given by interviewees available online, blogposts, or other articles written by interviewees).

Applying constructivist grounded theory

In the following section, I describe how I applied constructivist grounded theory in the three case studies (Yin 2014). In this chapter I only provide a general overview of how the collection and analysis of the data was conducted. A more detailed description can be found in Appendix A. The analysis of the data was conducted with the software HyperResearch in the first case study, and with the TAMS Analyzer (Weinstein 2006) in the second and third case study. Both HyperResearch and the TAMS Analyzer are designed for qualitative data analysis, and offer tools for qualitative coding, memo writing, and options to visualize and analyze the coding scheme. I also made intensive use of mind mapping software to explore dimensions and relationships between categories.

All three case studies roughly followed five steps:

1. Collecting an initial data sample

Following an extensive literature review, the first step in the process of constructivist grounded theory is collecting an initial data sample. To study data activism in the first two cases, I decided to focus on 'best practice' organizations early on – namely, the OKF DE and mySociety (see details in Chapter 1). The two criteria for selection were influence and embeddedness. Concerning influence, best practice organizations in civic tech shape the field in the sense that they provide orientation for others through the applications that they invent, and through the way that they communicate what those applications are about. As data activists have not been thoroughly researched

to date, best practice organizations can provide a useful starting point for the analysis. In addition, the focus of this study should be on organizations that are embedded within a broader international community, and take part in a continuous international exchange with other civic tech or open data organizations, both online and at international conferences. This type of embeddedness makes it unlikely that the studied organizations are extreme outliers. As explained in Chapter 1, the third case study about the entanglements between data journalists and data activists did not focus on any specific organization. Instead, I interviewed data journalists in different organizational settings to explore their relationship with data activists, e.g. national news media, startups, and non-profit newsrooms. The data collected for the previous case studies was taken into account to compare the practices and imaginaries of the different actors.

In each case study, I made use of the extensive content available online. For the first two case studies, where the focus was on specific organizations, I included official homepages, project-specific websites, and important self-portraying documents – e.g., the 'Open Definition' (Open Knowledge n.d.) for the OKF DE, and Crabtree (2003) for mySociety. Because both the OKF DE and mySociety had relatively small core-teams at the time the case studies were conducted (2013-2016), the goal was to interview as many relevant members as possible. I was primarily interested in interviewing developers and members in leading positions. Developers were important to my research because of my use of reconstruction interviews, while leading members (e.g. CEOs or project managers) were important because they articulate the social imaginaries and leading visions of the organization. In both cases, I first conducted two interviews to complement the online content.

For the third case study, the initial data sample was developed from my participation at the Daten Labor conference and a local meetup (#DDJ Berlin). Both events were attended by data journalists in almost all major national newspapers, and some local ones with 'data teams'. The notes and observations from those events guided an initial selection of interviewees. I collected available content online (such as interviews given by my interviewees elsewhere) to prepare the interviews, and many of these materials have been included in the post-interview analysis.

2. Initial coding

The initial data samples formed the basis for what Charmaz (2006) calls 'initial coding'. Put simply, coding in grounded theory means attaching labels to segments of data that reflect what each segment is about. These codes provide "an analytic handle to develop abstract ideas for interpreting each segment of data" (Charmaz 2006, 45). Through coding, the data is broken into segments and compared with other data, to gradually build levels of abstraction. Initial coding represents the first step in this analysis, where the codes are more descriptive and close to the data at hand. Charmaz (2006) suggests different techniques, such as line-by-line or incident-by-incident coding at this stage.

Given the focus on practices of this thesis, in all three case studies I initially coded primarily verb-by-verb. Because the work of the OKF DE and mySociety revolves around specific projects, I initially also coded project-by-project. In other words, I assigned verbs to projects. For example, interviewees used verbs like 'helping', 'annotating', 'deep linking', 'pushing', or 'scraping' when they spoke about mySociety's parliamentary monitoring website TheyWorkForYou. In the third case study, where I interviewed journalists, I place a stronger emphasis on how interviewees position themselves

professionally, and describe their relationship with data activists. I generally followed Charmaz's (2006, 49) advice to code for actions with gerunds to avoid applying pre-conceived concepts to the data (e.g. by coding with 'stating' rather than 'statement'). After I coded all the of material in this way (interviews, online content, field notes), I combined similar verbs into more abstract practices.

3. Focused coding

After the initial coding I employed what Charmaz (2006) calls 'focused coding', to move towards more analytic categories that abstract common themes and patterns from several codes. Focused coding "synthesize[s] and explain[s] larger segments of data" (Charmaz 2006, 57). The developed categories were tested across all of the collected materials, to examine their validity. Was the category applicable to the whole data, was it incisive, and did it have some explanatory power? The goal of focused coding is to determine the most significant categories and involves creating a hierarchy of categories and subcategories (Strauss and Corbin 1998).

Strauss and Corbin (1998) separate what Charmaz (2006) calls focused coding into 'axial' and 'selective' coding. Axial coding involves exploring the causes, consequences and contexts of categories. They suggest the use of paradigm models that systematically explore different dimensions and conditions of categories (Strauss and Corbin 1998). Charmaz (2006, 62) advises against a strict application of paradigm models, as this "may limit what and how researchers learn about their studied worlds and, thus, restricts the codes they construct", and suggests a more practical and cautious use. I did not conduct axial coding because I found the systematic application of a coding paradigm too restrictive and alienating. Instead, I followed the leads in my data and

concentrated on developing theoretical sensibility (see Charmaz 2006, 135), e.g. by looking at an issue or category from different angles, establishing connections, and asking questions. Helpful here were some of the techniques proposed by Strauss and Corbin (1998), e.g. the 'flip-flop' technique of imagining the extreme opposite of a category.

4. Theoretical sampling

In the logic of grounded theory, coding the initial data sample starts the 'spiral' of developing the theory (see discussion of grounded theory above). I developed preliminary and incomplete ideas and concepts through initial and focused coding, but I needed to collect more data to test these ideas and/or to find new leads. Following my initial sampling I conducted theoretical samplings, and selectively included new content, requested interviews, or attended workshops and conferences. For example, after the analysis of the initial sample during the second case study, I realized the importance of international work for mySociety. Because I only had a few general remarks about it, I contacted and interviewed a member of mySociety's international team in the second spiral to further explore this aspect and develop my categories. After the new data had been collected, I started a new spiral through the basic steps of constructivist grounded theory, i.e. I repeated initial coding on the new data and compared it with the old one, and I applied my previous categories to the new data, with the result that some of these categories were modified or dismissed and others became more central. In the first two case studies I conducted three spirals, i.e. I expanded my data following a theoretical sampling twice after the initial data sample had been analyzed.

For the third case study, the process of analysis and data collection guided by theoretical samplings was much more iterative and continuous due to my direct local involvement in Berlin (see Appendix A). I also included the material and the results of the first two case studies, in order to directly compare the perspectives of data journalists with those of data activists.

5. Finalizing the theory

Traditionally, grounded theories are considered 'completed' once 'theoretical saturation' is achieved. Theoretical saturation is usually defined as the state when gathering more data about a theoretical category "reveals no new properties nor yields any further theoretical insights" (Charmaz 2006, 189). The concept of theoretical saturation is problematic, because the term 'saturation' is ambiguous and implies that every variable has been explored in full, which was outside the scope of this thesis (see more in the evaluation below). In most cases it is more accurate to speak of 'theoretical sufficiency', as Dey (1999) suggests.

In the first two case studies, I achieved theoretical sufficiency with the development of a core category that integrates and elucidates the central aspects of my theory. This follows Strauss and Corbin's (1998) suggestion to find and develop one core category that helps to narrate the theory similar to a story, in a research process they describe as 'selective coding'. Their requirement of *one* core category is frequently criticized. Charmaz (2006) argues that while theories require structure, they should also be able to cope with ambiguity, and can have several core concepts that may not always easily align with each other. However, developing one core category with a narrative that was able to integrate all the other concepts was suitable for the first two case studies that I conducted, because they were focused on particular

organizations that similarly work on creating a narrative around their own history, mission, and self-understanding.

By contrast, the third case study did not result in one elusive core category, but in three distinct yet interconnected groups that exist along a shared continuum of practices and understandings, which oscillate between 'facilitating' and 'gatekeeping'. 'Facilitating' builds on the core category developed in the second case study, which illustrates how much the three case studies complement each other.

Evaluation

In this section, I discuss the deliberate limitations of the methodological approach used in the three case studies: the focus on describing social imaginaries rather than tracing their influence, and the cultural and historical situatedness of the theories I developed. In addition, I address some of the common particularities of grounded theory.

Limitations in scope

There are two main limitations in the methodological approach presented in this chapter. First, I provide an analysis of the social imaginaries and practices of data journalists and data activists because I *anticipate* that their practices and modes of understanding are influential, due to their status as pioneer communities. It is outside the scope of this thesis to empirically trace this influence, or to show the ways in which the practices and ideas of data journalists and data activists are adopted by others, e.g. by looking at how other civil society actors take inspiration from civic tech organizations. Second, I focus on exploring the entanglements between data journalists and data activists *today*, and not on how those entanglements slowly developed over

time. However, my analysis provides a useful starting point for subsequent studies, for example a historical analysis, or research into how the practices and social imaginaries examined in this thesis are being adapted elsewhere.

Concerning the study of data activism in Chapters 4-5, I consider the theoretical concepts presented here as starting points for studying data journalism and data activism more broadly. My theories are contextually situated in time, place, and culture because I concentrate on specific and unique organizations in Central and Northern Europe. While I am confident that I have covered these organizations adequately to answer my research questions, it is important to be aware that they do not represent civic tech or the open data movement as a whole. There are influential civic tech organizations in Latin America or Asia, for example, which presumably understand and enact civic tech in ways that differ from those presented in this dissertation. The quantitative network analysis of the connections between civic tech organizations on GitHub illustrates how much the civic tech scene is separated geographically, on a global scale (see above). Yet while they are not representative for the phenomenon as a whole, the organizations I examine in this thesis are influential internationally. Therefore, my analysis should provide a useful reference point for studying data journalism and data activism in other regions.

The identification of different groups within and beyond data journalism in Chapter 6 is similarly situated, but this situatedness was easier to mitigate in comparison to the study of data activists in Chapters 4-5 because there has been a lot of research on data journalism in different national contexts (see Chapter 2). Some of this research was useful in the analysis of the data gathered for the third case study. The theory developed here is compatible with, and extends upon, the existing research literature. However, there is room to

further develop the categories I present in Chapter 6. My data indicated, for example, that more distinctions can be made among groups of journalists who are working in startups than I present in Chapter 6. I did not further explore these differences because it was not necessary to sufficiently answer my research questions, and a further development was not possible due to time constraints. A follow-up study could concentrate on data journalists working in startups to develop a more differentiated picture, and further elaborate the continuum between facilitating and gatekeeping that I develop in this study.

The implementation of grounded theory: Relationship with participants and use of research literature

Because *constructivist* grounded theory insists that theories are social constructions that do not emerge from an objective, observable truth, it is important to reflect my own role as a researcher and my relationship with those I studied. In many ways the theories I present here are co-constructions, because they emphasize understanding rather than explanation; which means that I also learned a lot from my interviewees. In addition, my interviewees were highly self-aware and reflective about their own work. They developed their own theories about civic tech and data journalism and were aware of, and acknowledged, some of the common criticism around their work (e.g. that open data and civic tech would 'empower the empowered', see Chapter 1). They subsequently showed great interest in my research project and were willing to do interviews with me. Many of them asked me to share the results with them. Before publication, I therefore shared the articles with all of the interviewees and invited them to comment. Six interviewees replied. Their feedback was positive and confirmed my analysis, but also provided me with some interesting angles on how to interpret the results. This feedback

indicated the fit and relevance of my theory, which are two important characteristics for the quality of grounded theories (Charmaz 2006, 54).

Finally, a few notes about my usage of the research literature. Some grounded theorists insist that the researcher should avoid existing theory literature before the analysis, in order to be open-minded and avoid applying pre-conceived categories to the data. I think this idea is not only unrealistic but is problematic, as it tends to conceal the researcher's presuppositions and values. As Dey puts it: "There is a difference between an open mind and an empty head" (quoted in Charmaz 2006, 48). As long as existing concepts and theories are not blindly attached to the data, they can provide useful guidance to develop theories. Contrary to the classic stance of grounded theory, the aim of this thesis is to address gaps in the existing research literature identified after an extensive literature review (see Chapter 1 and 2). Rather than approaching the theory-construction with an 'empty head' and starting from zero, I tried to extend existing literature from the very beginning. Existing research also helped in the analysis of the data, though I tried to avoid relying on existing concepts during the initial coding phase. When my categories became more abstract and analytical, I compared them with existing concepts to check for overlaps or alternative explanations. Using this approach, I am better able to contribute to the existing research through my empirical and theoretical work – the results of which I present in the following.

4. Datafication and empowerment

How the open data movement re-articulates notions of democracy, participation, and journalism

Abstract: This article shows how activists in the open data movement re-articulate notions of democracy, participation, and journalism by applying practices and values from open source culture to the creation and use of data. Focusing on the Open Knowledge Foundation Germany and drawing from a combination of interviews and content analysis, it argues that this process leads activists to develop new rationalities around datafication that can support the agency of datafied publics. Three modulations of open source are identified: First, by regarding data as a prerequisite for generating knowledge, activists transform the sharing of source code to include the sharing of raw data. Sharing raw data should break the interpretative monopoly of governments and would allow people to make their own interpretation of data about public issues. Second, activists connect this idea to an open and flexible form of representative democracy by applying the open source model of participation to political participation. Third, activists acknowledge that intermediaries are necessary to make raw data accessible to the public. This leads them to an interest in transforming journalism to become an intermediary in this sense. At the same time, they try to act as intermediaries themselves and develop civic technologies to put their ideas into practice. The article concludes with suggesting that the practices and ideas of open data activists are relevant because they illustrate the connection between datafication and open source culture and help to understand how datafication might support the agency of publics and actors outside big government and big business.

Introduction

Agency is deeply connected to the distribution of knowledge and power. If we understand agency as "the longer processes of action based on reflection, giving an account of what one has done, even more basically, making sense of

the world *so as* to act within it" (Couldry 2014, 891), then the conditions under which we can make sense of our world and our own actions are crucial for our capacity to act with agency. With the expansion of the Internet, social media, and big data technologies, we can currently observe a number of fundamental transformations of knowledge production and distribution that raise urgent questions about public agency. To date, however, questions about agency have been "obscured by unnecessarily generalised readings" (Couldry and Powell 2014, 1) of the supposed power of the new technologies. For this reason, Couldry and Powell (2014) recently called for 'social analytics' as a new research paradigm in relation to big data.[14] They stress that agency is still relevant and that we should study new forms of reflexive agency in increasingly datafied societies. This involves paying more attention to the social and cultural dimension of this transformation by examining how social actors respond to processes of data collection and analysis and how they use data to "meet their own ends" (Couldry 2014, 892). This article follows Couldry and Powell's call and shows how activism around open data is a rich, but so far largely overlooked side of inquiry that allows us to think about the relationship between data and agency in new ways.

The open data movement is a particularly interesting case because it intersects with two ongoing transformations of knowledge and power that seem to contradict each other in terms of agency: datafication and the proliferation of hacking or open source culture. On the one hand, the practices and imaginaries of open data activists are centered around the distribution and use of data and

[14] The social analytics approach was first developed in the Storycircle project to study how organizations use analytics to meet their goals. See http://storycircle.co.uk/.

thus linked to datafication, the ubiquitous quantification of social life (Mayer-Schönberger and Cukier 2013, 78), for which big data is the most prominent expression. big data "reframes key questions about the constitution of knowledge" (boyd and Crawford 2012, 665) and raises concerns about the agency of publics. As Couldry and Powell (2014, 4) note, big data technologies and the growing relevance of algorithms may disconnect "system and experience" because the traces of data people leave behind are often unconscious and not meaningful to them, and the insights generated by companies or governments are not, or only partially, "folded back into the experience of everyday life". The comprehensive surveillance of online activities made possible by big data technologies thus might impede our potential to act in an agentic manner. On the other hand, open data activists apply practices and values from open source culture to the creation and use of data. This links them to other initiatives rooted in open source culture, like Open Access, Wikipedia, Wikileaks, Anonymous or Creative Commons (Beyer 2014b, 2014a; Coleman 2014; Sauter 2014). Similar to datafication, these phenomena raise fundamental questions about "the nature of knowledge and expertise, how information is organized and evaluated, and who decides" (Lievrouw 2011, 26). Different to datafication, however, open source culture is associated with a transparent and collaborative form of governance that might support agency. As Raymond (2001) famously pointed out when he contrasted the 'bazaar model' of open source with the 'cathedral approach', open source culture is fundamentally concerned with the rights to access and distribute knowledge. Open source is based on voluntary participation (Weber 2004, 62) and collaboration, granting access to the source code of software and incorporating contributions from potentially everyone. The implications of transferring the "open source process" (Weber 2004, 16) and the values

inherent in this process to new domains with different ways of organizing knowledge "reach directly into the heart of the legitimacy, certainty, reliability and especially the finality and temporality of the knowledge and infrastructures we collectively create" (Kelty 2008, 6–7).

An analysis of the open data movement offers a unique opportunity to *connect* datafication and open source culture, which raises interesting questions about agency: how do activists apply practices and values from open source culture to data, and what does this tell us, in return, about agency in datafied publics? To address these questions, I will present key findings from a study on the *Open Knowledge Foundation Germany* (OKF DE), a not-for-profit organization and one of the most influential and visible actors in the German open data movement.[15] First, I will address how we can trace the influence of open source culture on open data activists. This will be the foundation for the following analysis, in which I will show how open data activists modulate open source practices by applying them to data. Inspired by the social analytics approach, this analysis will ask, first, what their practices and imaginaries tell us about the *conditions* under which datafication might support the agency of publics and, second, how datafication supports the agency of activists themselves. In the conclusion, I will reflect on the broader relevance of this type of activism for the ongoing datafication of social life.

[15] http://okfn.de/. The OKF DE was founded in 2011 as the first international chapter of the British Open Knowledge Foundation, which was founded 2004 in Cambridge. It was recently renamed into 'Open Knowledge'. See https://okfn.org/.

Tracing the influence of open source culture on open data activists

It is generally acknowledged that activism around open data is rooted in hacking culture (cf. Bates 2012; Davies 2010; Johnson 2014), or more specifically open source culture as one of the most prominent genres of hacking (Coleman and Golub 2008). However, while this connection is frequently pointed out, it is rarely examined in more detail. Authors usually refer to a set of broad ethical commitments taken directly from traditional hacker culture. These ethical commitments have been famously described by Levy (1984) as follows: access to computer technology and information should be free, centralized forms of power are rejected in favor of decentralization, hackers adhere to a meritocratic culture of technological excellence in which the hacker should only be judged by his or her code, and the belief that computers can create a 'better world'. While these principles are indeed relevant, we run the risk of oversimplifying the relationship between open data activism and open source culture if we solely rely on them. As Coleman (2013, 17) points out, the frequent reference to Levy's account is problematic because it "whitewashes" the diversity among hackers. While hackers share some technical and ethical commitments (for which Levy's description is still useful), hacker culture should not be treated as a "singular code formulated by some homogeneous group called hackers but instead as a composite of distinct yet connected moral genres" (Coleman 2013: 19).

To develop a more nuanced and differentiated picture of how open data activists draw from open source culture, we can turn to research on its broader cultural significance and influence beyond software development. Particularly helpful here is an approach developed by Kelty (2008). While most attempts

to grasp this phenomenon are primarily interested in making generalizations – for example by asking whether diverse initiatives rooted in open source culture are forming a coherent movement with a political project (cf. Beyer 2014a; Clement and Hurrell 2008; Kapczynski 2010) or how the organizational features of open source software development can be generalized and applied to new domains (cf. Demil and Lecocq 2006; Matei and Irimia 2014; Weber 2004) – Kelty developed a model that can be used to trace the influence of open source culture for specific cases. In his study of the cultural significance of free software,[16] Kelty suggests that open source advocates associate with each other not just through a set of ethical commitments, but through a range of key practices and social imaginaries (Taylor 2004). He understands open source as an experimental system made up of five key practices or 'components': sharing source code, defining openness, writing copyright licenses, coordinating collaborations, and forming a movement. Understood in this way, open source becomes "a system of thresholds, not of classification" (Kelty 2008, 16):

> Within each component are a range of differences in practice, from conventional to experimental. At the center, so to speak, are the most common and accepted versions of a practice; at the edges are more unusual or controversial versions. (Kelty 2008, 15)

Due to their flexibility, these components are not exclusive to the development of software: each of these practices can be adapted or

[16] 'Free software' and 'open source' generally refer to the same practices. While the term free software emphasizes social and cultural values ('free as in speech'), open source emphasizes the practical advantages for developing software (Kelty 2008). For convenience, I will use 'open source' to address both strands.

'modulated' to apply them to other domains. Therefore, Kelty (2008, 246) calls initiatives like Wikipedia or Creative Commons *modulations* of open source that emerge "out of a direct engagement with and exploration of Free Software" and are "committed to experimenting with the given practices of Free Software". Creative Commons, for example, paralleled some of the work of the Free Software Foundation in a different context (Garcelon 2009, 1315): it modulates the practice of 'sharing source code' by applying it to 'content', it is writing copyright licenses for this new type of 'source code', and it has become a movement as well.

To trace the influence of open source culture, Kelty (2008, 278) suggests treating its key practices as a template that interacts with other forms of knowledge management: "Where the practices match, no change occurs, and where they don't, it is the reorientation of knowledge and power". Therefore, the proliferation of open source culture can be described as the proliferation and modulation of its key practices in order to alter the means of knowledge production and circulation. Tracing the influence of open source culture on open data activists then comes down to a set of specific questions: Which practices are modulated? How are they modulated? How does this change the domain to which they are applied? Answering these questions will help us to grasp how activists try to apply the more transparent and collaborative forms of governance associated with open source to politics, and how this might support the agency of datafied publics.

Practices and imaginaries of open data activists[17]

The following analysis is based on 10 semi-structured interviews with members of the OKF DE core team (including the chairman and founder, main developers, committee members and project managers) and a content analysis of nine relevant documents that were selected using a theoretical sampling, for example self-portraying descriptions from the official homepage. The data was collected in three rounds between September 2012 and January 2013 and analyzed using a grounded theory approach (Glaser and Strauss 1967). I will structure the presentation of the findings in a way that shows how one modulation of open source culture leads activists to other, subsequent adaptations and interpretations.

1. By regarding data as a prerequisite for generating knowledge, activists transform the sharing of source code to include the sharing of raw data. Sharing raw data would allow others to make their own interpretation of it and generate their own knowledge, which represents a 'democratization of information' for activists.

2. Seeing information as a necessary precondition for political participation, activists connect this idea to an open and flexible form of representative democracy by applying the open source model of participation (the 'bazaar model') to political participation, which should lead to more

[17] Some of the empirical findings presented here were published in German by the author (Baack 2013).

participation of citizens in political decision-making processes and more active and engaged local communities.

3. A third set of practices refers to activists' acknowledgment that raw data needs to be 'refined' to create knowledge for citizens, which is why they seek to create, and become, 'data intermediaries' for the public. This leads them to a special interest in journalism.

In the following, I explain each of these modulations and their implications in more detail.

Raw data as source code

The overall mission of the *Open Knowledge* Foundation is already implicated in its name. The organization adapts a hierarchical understanding of the relationship between data, information, and knowledge that is common in knowledge management literature (Tuomi 1999). According to this model, (raw) data are understood as symbols that have not been interpreted; data becomes information when it is structured and put into context; and information becomes knowledge when it is interpreted, meaningful, and actionable. As the OKF explains, open knowledge "*is what open data becomes when it's useful, usable and used*" (Open Knowledge n.d.). With its name, the OKF thus indicates that it aims at spreading not just open data, but open knowledge. However, data is seen as a *prerequisite* for generating knowledge. This hierarchical understanding leads activists to the first and most fundamental modulation of open source culture: to conceive raw data as source code that should be shared openly to allow others to interpret it and to generate their own knowledge from it.

Implicit here is that activists do not simply modulate the practice of sharing source code by replacing code with data. They also adapt the metaphors and concepts behind this practice. To execute human-readable source code on a computer, it has to be translated into binary instructions that are only readable by machines. These binary instructions cannot be retranslated back into the source code from which they have been generated. Having only the binary code without the source code (which is the case for most proprietary software) means that it is not possible to understand or modify the 'inner workings' of the software. Similarly, open data activists treat raw data as source code and interpretations – or knowledge – as binary code. As one activist explains, raw data "is not really neutral" but it allows more interpretation than a "summary or a press conference" (Interview: Developer 1).[18] That is because summaries are *already* interpretations of raw data. Only offering an interpretation of raw data without allowing access to it would make it difficult for others to understand how this interpretation was developed. Governments would then maintain a 'monopoly of interpretation'. Sharing raw data makes the process of interpreting it transparent and breaks governments' monopoly, which means that everybody could make his or her own interpretation of the data that governments use to make and justify their decisions – allowing people to examine biases in government's data collection and interpretation. For activists, open data therefore represents a democratization of interpretation or – as they put it – a "democratization of information" (Interview: Chairman & Founder).

[18] Quotes from the interviews were translated from German by the author.

It is interesting to contrast the notion of 'raw data' developed by activists with the way the term is used in discussions about big data. "Raw data is an oxymoron" (Gitelman 2013) is one of the most common critiques of big data advocates' belief in "objective quantification" (van Dijck 2014, 198) or big data's "aura of truth, objectivity, and accuracy" (boyd and Crawford 2012, 664). In their critique, authors point out that data is always prefigured through gathering mechanisms (van Dijck and Poell 2013, 10) and collected data has to be interpreted to make it meaningful and actionable, a process guided by specific interests and rationalities and not something that can be considered as objective. Essentially, this questions whether something like 'raw' data actually exists when we understand it as something 'pure' beyond human influence. However, members of the OKF DE adapt a different understanding of 'raw' data. For them, 'raw' simply means 'as collected'. Accordingly, sharing data in 'raw' form – 'as collected' – is not about revealing an unbiased and objective truth, but about making the biases of this data *transparent* and allowing "more interpretation of truth" (Interview: Chairman & Founder).

Using this understanding as a basis, members of the OKF DE are also concerned with the conditions that must be met to ensure this type of transparency, i.e. with the way raw data has to be provided to fulfill their vision of a democratization of information. This leads them to another modulation of open source practices: defining openness. More specifically, they define both the legal and technical characteristics of openness in relation to data in order to delineate open data from 'closed' data. For legal openness, the OKF developed the international Open Definition (Open Knowledge n.d.), according to which data is 'open' when it can be accessed, modified and shared by anyone for any purpose without restrictions. Technical openness is about ensuring that these rights can be exercised without too much effort. Key here

are the principles developed by the Sunlight Foundation (2010) and the rating system developed by Tim Berners-Lee (2010). According to these guidelines, datasets should be complete, released in a timely fashion, accessible, machine readable, and available in open formats. While activists acknowledge that personal data and data crucial to security should not be made available in this way, they suggest that these legal and technical conditions are necessary to effectively break the interpretative monopoly of governments.

Given the importance of knowledge for agency, this type of transparency has the potential to support the agency of datafied publics. As activists acknowledge themselves, however, the mere provision of raw data is insufficient and only represents "the first step" (Interview: Chairman & Founder). As I will explain in the following sections, this provision should go along with more continuous and flexible forms of participation and 'data intermediaries' that make raw data accessible to the public.

Data and democracy

The democratization of information described above is not regarded as an end in itself by activists. Ultimately, this form of transparency is taken as a means through which "the people should be considered again as the sovereign" (Interview: Project Manager 1). Even though they do not explicitly talk about agency themselves, activists' articulations of their broader aims are interesting for understanding *how* and under what *conditions* the democratization of information they envision could support the agency of datafied publics. The overall 'vision' of OKF DE members is essentially a vision of citizen empowerment: sharing raw data should help citizens to better understand and control their governments and to be more active and engaged in their local communities.

This means that more possibilities for citizens to participate in political decision-making processes is a major goal for members of the OKF DE: "to participate, people need information" (Interview: Project Manager 2). In this respect, they regard themselves as part of an "Internet generation" that is not content with periodic voting: "[I want] a higher degree of participation (…) a more continuous form of participation" (Interview: Committee member). This does not, however, necessarily translate into a demand for more direct democracy. Instead, the open source model of participation is taken as a paradigm:

> What is powerful about open source development is that people can elect themselves as participants. I mean people can find my project and then decide for themselves to participate in its development and contribute to it. I think this model of self-selective participation is extremely powerful and I believe it can be applied to politics. (Interview: Developer 1)

This means that everybody who wants to participate in the decision-making process of a particular issue should have the opportunity to do so in a meaningful way. Here, activists explicitly modulate another practice from open source culture: coordinating collaborations, the organization of open source projects (Kelty 2008, Chapter 7). As mentioned above, this organizational model has been described as the 'bazaar model' (Raymond 2001) because it encourages and incorporates contributions from potentially everyone. Just as there is not one standard model for coordinating collaborations in open source – larger and more well-known projects like the Linux kernel, the Apache servers, or the Debian project have all developed distinct organizational models over time (Kelty 2008; Coleman 2013; Weber 2004) – activists reject clearly prescribing a specific model of participation. For them, applying the bazaar model of open source to governance is first and foremost about

experimentation. There will not be "this one solution that you just need to apply. I think public authorities will need to have the courage to experiment" (Interview: Chairman & Founder). This illustrates that more participation is not seen as a natural outcome of open data. Activists argue that it requires a cultural change within public institutions: a change towards a "beta culture" that is willing to experiment and risk failure (Schwegmann 2012), and a more collaborative and less authoritative relationship with citizens. Public institutions, it is argued, should promote the use of data and actively include citizens in decision-making processes: "It is not just about opening data (…) but also about investments from public institutions to ensure that this data is used" (Interview: Committee member). Activists think that this cultural change will mainly happen at a local level, where issues are "closer" to the people and institutions can experiment with "less resources" (Interview: Chairman & Founder).

Taken together, the way activists apply the open source model of participation to governance results in a notion of a more open and flexible form of representative democracy. 'Open' refers to a higher degree of transparency (by sharing raw data) and the openness of political decision-making processes for public participation. 'Flexible' means that activists think that the inclusion and coordination of citizens' voluntary, 'self-selective participation' should be adapted to the issue at hand and to the local context. At the same time, activists do not question representative democracy as such and are rather skeptical about elements of direct democracy, e.g. referendums: "I don't know if direct democracy is always the right answer (…) but I definitely want more mechanisms to involve people more often" (Interview: Chairman & Founder). From the perspective of democratic theory, they negotiate between representative models of democracy – in which participation is mainly limited

to periodic voting – and direct models of democracy, where entire electorates vote on certain proposals. This is similar to Barber's (2004) model of 'strong democracy', a more explicit attempt to develop an alternative to representative and direct democracy. Put briefly, strong democracy is based on a "creative consensus" that is meant to recognize the diversity of interests and "is premised on citizens' active and perennial participation in the transformation of conflict through the creation of common consciousness and political judgment" (Barber 2004, 224). Similarly, the diverse and flexible modes of organizing voluntary participation envisioned by OKF DE members require the active involvement of citizens and imply a consensus building process that is 'creative' in negotiating diverse interests *and* in its organization.

We can summarize the ideas and aims of open data activists described thus far to articulate – as an intermediary result – a first proposal about the conditions that must be met to support the agency of datafied publics: the transparency created through the sharing of raw data should be accompanied by a cultural change within public institutions to support voluntary and flexible forms of participation similar to those found in open source projects. As I will detail in the next section, activists not only emphasize the importance of public institutions, but also of other intermediaries to facilitate this participation.

Creating empowering intermediaries: Complementing or replacing journalism?

Even though the idea behind the democratization of information is to *potentially* allow *everybody* to interpret raw data, activists are well aware that the average citizen does not have the time and expert knowledge to do so. They recognize that their vision of empowerment through open data can only be

realized with intermediaries that make raw data accessible to the public.[19] Different to the modulations discussed above, this is not directly reflected by open source culture. However, it emerged out of activists' engagement, exploration, and modulation of open source, i.e. of modulating the sharing of source code to include the sharing of raw data and of using the open source model of participation as a paradigm for political participation. Because activists realize the importance of suitable intermediaries for their goals, they actively seek to 'create' them – which makes this aspect also interesting for understanding how datafication supports the agency of activists themselves.

In terms of agency, more interesting than the basic acknowledgment that intermediaries are necessary is what kind of intermediaries are deemed necessary to empower citizens. Three criteria can be identified that constitute an 'empowering intermediary' in the eyes of activists. First, they should be *data-driven*, which means that they should be able to handle large and complex datasets to make them accessible to others. Second, empowering intermediaries should be *open*, which means that they should make the data from which they generate stories or build applications available to their audiences – the principle of sharing raw data applies here as well. The fact that professional journalists or NGOs often do not give access to their sources is therefore frequently criticized, one activist calling it a "fundamental bug of newspapers" (Interview: Developer 1). Third, empowering intermediaries should be *engaging*, which means that they should actively involve citizens in

[19] I adapt activists' usage of the term intermediary in this article. However, 'intermediaries' in their sense are more likely to act as mediators according to Latour's (2005, 39) distinction of the terms.

public issues. This implies that such intermediaries should not only be information providers and that they should have a cooperative relationship with their audiences: "journalism also needs to change to be closer to citizens" (Interview: Developer 2). Taken together, the three criteria articulated by activists are clearly related to their goal of an open and flexible democracy described above. Empowering intermediaries are a necessary prerequisite in this sense, or, as we might say in relation to agency, they are important supporters of agency in datafied publics.

To 'create' these intermediaries, activists try to cooperate with other NGOs and professional journalists and offer teaching. Here, they are part of a larger phenomenon: the increased interaction between the social worlds of technology and journalism, or more specifically between hackers and journalists (Lewis and Usher 2013; Parasie and Dagiral 2013; Karlsen and Stavelin 2014; Royal 2010). Members of the OKF DE are involved, for example, in Hacks/Hackers events (Lewis and Usher 2014), where hackers and journalists come together to innovate news; the News Challenge of the Knight Foundation, an open-to-all contest rewarding projects that aim to transform news and information distribution (Lewis 2012a); or the Knight-Mozilla Fellowships, which bring together hackers and technologists "to spend 10 months working on open source code with partner newsroom[s]" like *The New York Times* or *Der Spiegel* (OpenNews n.d.). The existing research on the interaction between these groups shows that activists' goal of a more equal and cooperative relationship between citizens and professional journalists essentially questions the professional boundaries of journalism. As Lewis (2012b) describes, journalism is shaped by a professional logic of (exclusive) control over content that does not align easily with a more participatory form of journalism favored by activists. The research shows an ongoing process of

negotiation of different values, imaginations, and practices: news organizations try to re-interpret the distinct way activists think about technology and data "into the language of news" (Lewis and Usher 2013, 604), while at least some activists tend to believe the new possibilities are "capable of altering the very nature of journalism" (Parasie 2011).

However, activists do not only try to influence journalism by interacting with professional journalists or by becoming programmer-journalists in newsrooms. They also act as intermediaries outside the profession and develop independent, non-profit applications to 'implement' their ideas. Key here are so called 'civic technologies' – small-scale, specialized applications that aim to "connect people" (Interview: Developer 1). These applications are either about improving government services for citizens, or about helping citizens to coordinate with each other to solve problems together. Often these are relatively simple web applications that focus on one task. For example, there are civic technologies that help people to exchange deposit bottles, that show how and where to engage in local building projects, that inform people about the local air quality, visualize which parts of the city are barrier-free and which are not, and so forth.[20] Even though civic technologies do not always depend on open data, data is key to their functioning in two ways: first, the availability of open data creates more opportunities to develop civic technologies (for example, when they require traffic data); second, they often datafy the activities they are concerned with, i.e. they often create new data. For example,

[20] See a list of projects supported by the OKF DE at http://codefor.de/projekte/.

FragDenStaat.de (inspired by the British WhatDoTheyKnow[21]) makes it easier to submit freedom of information requests to public authorities and tracks both the requests and the responses from institutions. This crowdsourcing approach created a database that can be used to analyze and compare how different institutions react to these requests, what kind of requests are more likely to get refused and so forth. This illustrates that the development of civic technologies is not only interesting because it could support the agency of citizens. It also shows how activists use or create data to meet their own ends by developing tools to put their ideas into practice. For OKF DE members, the purpose of these applications is two-fold. On the one hand, they are supposed to help citizens to be more active and engaged in their local communities in a general sense – for example by helping people with disabilities to move around the city. On the other hand, they hope to create new communities or "alternative publics (…) with a controlling function" (Interview: Developer 1). An often cited example is Ushahidi,[22] which was originally developed by a group of citizen journalists to track violent outbreaks after a disputed election in Kenya (Giridharadas 2010). Because journalists received threats about their work, Ushahidi was designed as a crowdsourcing application that maps incidents reported anonymously by users. Both in the sense of more active and engaged citizens and of 'controlling publics', civic technologies are linked to a notion of "self-empowerment" (Interview: Chairman & Founder) or 'do-it-yourself-empowerment' through data,

[21] https://www.whatdotheyknow.com/.

[22] http://www.ushahidi.com/.

understood as the ability of citizens to solve issues without the help of governments or businesses.

In terms of agency, the development of civic technologies by activists is interesting for another, less obvious reason. Civic technologies can be described as alternative ways of fulfilling functions traditionally described as 'journalistic' (making governments more transparent and accountable and engaging citizens in public issues) or of accessing and using public services (e.g. with an easy-to-use website to submit freedom of information requests). In other words, these applications are developed *independently outside* professional journalism or public institutions, but at the same time are trying to fulfill *similar functions*. This ability to create 'actually existing alternatives' is characteristic of the political power of hacking in general, as Kelty (2008) illustrates. He argues that open source advocates (and modulations like the open data movement) create independent 'recursive publics' through their key practices:

> A recursive public is *a public that is vitally concerned with the material and practical maintenance and modification of the technical, legal, practical, and conceptual means of its own existence as a public; it is a collective independent of other forms of constituted power and is capable of speaking to existing forms of power through the production of actually existing alternatives*. (Kelty 2008, 3)

By being able to maintain their own terms of existence (to a certain degree at least), recursive publics can act as 'actually existing alternatives'. In this sense, civic technologies developed by activists could to some degree act as 'actually existing alternatives' to professional journalism or (ways of accessing) public services. Activists are well aware of this potential: The ultimate goal of developing alternative services with civic technologies is to pressure established institutions to adapt them. "Flagship projects" (Interview:

Chairman & Founder) are intended to demonstrate what is possible and to invite (or provoke) established institutions to imitate them. As one member notes: "We have discovered software as a lobbying tool" (Interview: Developer 2). Let me illustrate this with another example: Frankfurt-Gestalten.de (~'Shaping-Frankfurt') monitors information provided by local parliaments in the city of Frankfurt and illustrates them on a map. Users can check what is currently discussed in their street or district (e.g. building projects), comment on it or initiate new discussions. Activists use this project to advocate for easier access to local parliamentary data, and for local public institutions to offer similar services. Moreover, I suggest that applications like Frankfurt-Gestalten.de represent a data-driven form of local journalism that is focused on engaging citizens on a local level. As such, Frankfurt-Gestalten.de has a complex relationship with professional journalism: First, it could complement professional journalism because local journalists can use it as a research tool. Secondly, however, it also represents a potential threat for professional local journalism – if people use an application like Frankfurt-Gestalten.de instead of consulting their local news media. Yet it is also conceivable, thirdly, that news media develop and maintain similar applications themselves, offering them as services to their audience and using them as research tools for their own investigations – Bell (2014) recently made a similar suggestion. This example illustrates how activists attempt to directly or indirectly influence established institutions on many different levels through the development of civic technologies, and shows that acting as intermediaries themselves is as much about directly putting ideas into practice as it is about transforming existing institutions. It not only shows how activists use data to directly meet their own ends, but also how they attempt to influence the conditions of the wider public to support the agency of ordinary citizens.

Conclusion: Data hacking and new forms of agency?

I conclude by returning to the questions raised in the beginning of this article. What do the practices and values developed by members of the OKF DE tell us about the conditions under which datafication can support agency?

When we look at activists themselves, datafication obviously does not undermine, but rather supports their agency in important ways: their technological expertise enables them to utilize or create data to meet their own ends. They even use the applications they create as lobbying tools that pressure institutions by offering actually existing alternatives. These findings emphasize the connection between datafication and the proliferation of hacking culture. The ability to 'hack' and to create recursive publics fundamentally depends on the availability and modifiability of the underlying technology (Kelty 2008, 10–11): participants have to be able to access and modify the technology needed to build their own, independent infrastructures. Otherwise, the expressive use of technology – the expression of imaginaries, values and rationalities *through* technology – would not be possible. The process of 'datafying' a phenomenon – of transforming it into quantifiable information – can be an integral part of recursive publics in itself, as illustrated by civic technologies that collect data via crowdsourcing. More importantly, to datafy a phenomenon is to re-materialize it into a highly modifiable form: in its essence, data is structured information that can be analyzed, edited, and combined with other data. This is why the availability of data creates more opportunities for the development of software that utilizes it in new ways, like activists do with the civic technologies they create. By rendering phenomena into data that have never been quantified before, datafication can make the key practices of recursive

publics applicable to them in ways that have not been possible before – given that the data created in these processes is accessible and modifiable. Therefore, datafication also has the potential to extend both the possibilities for and the scope of recursive publics, i.e. of creating 'actually existing alternatives' to established forms of knowledge production and circulation.

Moreover, members of the OKF DE are primarily concerned with how they can support democratic values and the agency of citizens through open data. As I showed in this article, three interrelated conditions must be met in their eyes: raw data should be shared openly to make decision-making processes more transparent, public institutions should actively include citizens in these decision-making processes to create a more open and flexible form of representative democracy, and 'empowering intermediaries' are needed to make raw data accessible to the wider public. It seems clear that these propositions have a *potential* to remedy the issues identified by Couldry and Powell (2014), i.e. the danger that big data technologies undermine agency by disconnecting system and experience. However, it is of course important and necessary to critically examine these ideas. For example, how to ensure that the raw data provided by governments does not violate privacy and is free of manipulations? Can voluntary participation of citizens work at larger scales? How to ensure that these processes do not end up "empowering the empowered" (Gurstein 2011)? We also need to be wary about the idealism of activists and the high level of technical literacy required to get involved in data activism. It might be easy to criticize activists as naive or techno-deterministic when we point out all these potential issues. However, while there is much to be critical of, I suggest that it is equally important to study and understand the practices and ideas of activists in order to evaluate what we can learn from them. Activists aim to develop new rationalities and alternative social

imaginaries around datafication to *connect* system and experience in new ways and to create a new sense for the legitimacy of collective knowledge creation and distribution in democratic, datafied publics. For all their shortcomings, these attempts are relevant and deserve attention because they provide a vital starting point to discuss how we can counter the threats of big data and utilize the potential of these new technologies in ways that do not damage democratic values and the agency of those not in big government or big business. As Couldry and Powell (Couldry and Powell 2014, 4–5) point out, we should not only highlight the risks of creating and sharing data, but also the opportunities for forms of social organization that take "into account agents' practices of giving an account of themselves and their conditions of life".

A guidance for future research provided by this analysis is to look at the way activists' practices and ideas are institutionalized, i.e. how they are adapted by other NGOs, news media, or public institutions. As activists acknowledge themselves with their emphasis on the importance of empowering intermediaries, their influence on the wider public – and therefore their potential to support the agency of datafied publics – depends on transforming existing institutions rather than on building new, alternative ones. To study these processes, we can take further inspiration from Couldry's (2010, 1) concept of "effective voice" – the insurance that 'my voice matters', which is a crucial aspect for both agency and democratic legitimacy. We can argue that activists describe important preconditions for processes of effective voice in datafied societies. Yet we have to be critical about whether the adaption of activists' practices and ideas really leads to *effective* voice, or only to more opportunities to *raise* voice. What matters is how "people's practices of voice are *sustained* and the outcomes of those practices *validated*" (Couldry 2010, 113). We have to ask whether and how the adaption of activists' practices and ideas

by news media, public institutions, or others does or does not lead to structures that create and sustain the conditions necessary for effective voice in increasingly datafied societies. Such research can form the basis to further examine, refine, and extend the practices and imaginaries of activists to formulate in more detail the conditions necessary to support agency in the 'age of big data'.

5. Civic tech at mySociety

How the imagined affordances of data shape data activism

Abstract: This article examines how civic technologists understand and use data to facilitate civic engagement. It illustrates the cultural and historical situatedness of data affordances and shows how civic technologists think of themselves as facilitators of civic engagement: they use and understand data in ways that are supposed to expand the agency of publics towards governments. The article shows how civic technologists use data to implement a form of participatory culture in new ways, and contributes to our understanding of how datafication enables new forms of activism.

Introduction

The progressive datafication of social life is often perceived as a threat to democratic publics. Critics warn that filter bubbles would undermine dialog and consensus, that social discrimination will be reinforced (Barocas and Selbst 2016) or that 'surveillance capitalism' is fundamentally anti-democratic (Zuboff 2015). Yet as "democratic power is calculated power" (Rose 1999, 200), datafication is *also* driven by democratic visions and closely linked to notions of political accountability, fairness, and citizen empowerment. Forms of datafication driven by commercial players or governments have always been accompanied by civil society actors or journalists who have utilized data and related forms of quantification to advance their goals. However, given the new structures of (data) power that shape the workings of governments and businesses today, Milan and Van der Velden (2016, 6) suggest that it is increasingly important to investigate "how activism evolves in relation to big data". They argue that new forms of 'data activism' are "enabled and

constrained by data (…) and this special relation shapes tactics, identities, and modes of organizing" (Milan and Van der Velden 2016, 3).

In this article, I explore ways in which activism is enabled by datafication by looking at a group of actors who not only reacted to processes of datafication, but pro-actively embraced them: civic technologists. Civic tech is an umbrella term for diverse projects that attempt to make engagement easier for citizens, improve communication and feedback between governments and citizens, and strengthen political accountability. Among other things, civic technologists develop parliamentary monitoring websites, tools to help citizens report local infrastructure problems to local government, or freedom of information (FOI) websites that help users to submit freedom of information requests to public institutions. In its modern incarnation, civic tech is the result of a convergence between "communities of technological and political openness" (Yu and Robinson 2012, 195). Early examples include the British FaxYourMP (2003), which helped citizens to find and contact their representatives in UK parliaments (Townend 2008), or the monitoring website GovTrack.us (2004), which made information provided by the US Congress more accessible (Yu and Robinson 2012). From these early volunteer experiments, civic tech has grown substantially in recent years as it has been embraced by governments, corporations and foundations (Baraniuk 2013).

In addition to their success and growth, civic technologists are relevant because they act as pioneers for the use of data to facilitate civic engagement. While there have been predecessors of the tools they developed, the way they utilized data to make them accessible and offer additional services was novel. They took information that was available elsewhere and made it machine-readable, shared it openly, and built services on top of it, e.g. the ability to type in one's post-code to find one's representative in the British parliament on

FaxYourMP. Through such practices, they combined concepts of legal and technical openness in new ways (cf. Yu and Robinson 2012). Being pioneers of this type of work is also part of the self-conception of at least some civic tech organizations. They possess a sense of mission and make teaching their (data) skills and experiences to other civil society or media organizations an important part of their work.[23] Finally, non-profit civic tech organizations such as the Sunlight Foundation in the US or mySociety in the UK were also among the first to advocate for open data policies, and supported related freedom of information policies (Schrock 2016).

However, despite civic tech's success and potential influence, it has received little attention in media and communication studies to date. Most of the research that does exist is interested in how the phenomenon might reinforce existing power structures (cf. Gregg 2015), or who is using civic tech applications and in what way (cf. Cantijoch, Galandini, and Gibson 2016). What is missing is a nuanced understanding of the practices, ideas and motivations that guide civic technologists themselves and how those practices and ideas provide orientation for others. Critically examining civic technologists themselves is relevant because the broader impact of civic tech goes beyond the impact of individual civic tech applications. As Hepp (2016, 919) describes, pioneer communities such as civic technologists are influential in the sense that they develop "a horizon of *possibility* to which the everyday media appropriation of others orients itself, or at least can do so" (Hepp 2016, 919). Accordingly, we also need to be sensitive to how the practices and

[23] An example are the 'Schools of data' organized by *Open Knowledge*, which are active in various countries.

imaginaries civic technologists develop provide orientation for others, i.e. how other actors adapt and modify them.

This paper therefore critically examines how civic technologists understand and use data to "meet their social ends" (Couldry and Powell 2014, 2). What are the key practices of civic technologists in relation to data and how do they themselves understand what they are doing? Addressing these questions will contribute to our understanding of how activism is enabled or constrained by datafication, as it provides a basis for subsequent studies to examine if and how these practices and imaginaries can be found elsewhere, and how they might have been modified.

I present findings from a case study about the British non-profit organization mySociety. Founded in 2003, mySociety is one of the oldest and most influential civic tech organizations and arguably represents 'best practice' in the extremely diverse civic tech sector. Its UK websites have millions of users (mySociety 2015) and the organization had a direct influence on British policy-making (cf. www.parliament.uk 2014). Some of its more popular projects include FixMyStreet, which lets citizens report local problems like broken streetlights or potholes to local government; its right-to-know website WhatDoTheyKnow which helps users to submit FOI requests to public institutions; or its parliamentary monitoring website TheyWorkForYou which gives detailed information about voting records and makes parliamentary speeches more accessible. mySociety's projects are also prominent internationally and customized versions of its tools are used in 44 different countries (mySociety 2015), which let mySociety to transition from a UK-centric to an international organization. While it is not representative of the phenomenon as a whole, mySociety's success and international influence

provide a good starting point for studying the values and practices which shape civic technologists' use of data.

In the following, I shall first discuss my methodological approach. I follow a practice theory approach and use the concept of 'imagined affordances' (Nagy and Neff 2015) as a lens which helps amplify how members of mySociety themselves understand data and how they imagine it to advance their agenda. The remainder of the article describes how the data practices of mySociety relate to their broader imaginaries. In the conclusion, I will reflect on the implications for studying civic tech and data activism in general.

Researching the imagined affordances of data

To examine how members of mySociety understand and use data to meet their own ends, this paper relies on the concept of 'imagined affordances' (Nagy and Neff 2015). According to Nagy and Neff (2015), whatever actions a particular technology enables or constrains does not solely depend on its features or its material properties, but also on the *perception* of users and designers. Both may have "expectations about their communication technologies, data, and media that, in effect and practice, shape how they approach them and what actions they think are suggested" (Nagy and Neff 2015, 5). Applied to the subject of this paper, what data enables civic technologists to do does not solely depend on the properties of the data they collect or re-use, or on the applications they are able to develop with it; it also depends on how civic technologists themselves understand and perceive how data can serve their agenda. These perceptions and understandings are the basis for the "horizon of *possibility*" (Hepp 2016, 919) they develop as a pioneer

community, which affects the perceptions of other actors concerning how data can be used to facilitate forms of civic engagement and activism.

To study the imagined affordances of data for members of mySociety, this paper relied on a methodological approach inspired by practice theory. First, a focus on practices aligns well with the concept of imagined affordances because affordances enable or constrain certain actions, and people make sense of affordances "in and through practices" (McVeigh-Schultz and Baym 2015, 2). Second, practices are useful for examining the role of pioneer communities because they *act* as exemplars. This means that they not only communicate ideas and visions about how a technology can be used, but they also become influential because they demonstrate their own visions and thereby affect the perceptions of others. The practices they develop are expressions of their broader visions and we have to consider them inseparable if we want to understand their influence.

Accordingly, I employed methods that helped exploring what members of mySociety are doing and how they themselves understand and categorize what they are doing in relation to data (cf. Couldry 2004, 2012). I followed a constructivist grounded theory approach (Charmaz 2006). Grounded theory was chosen because of its core principle of theoretical sampling, i.e. an initial data sample is continuously expanded with new data to systematically elaborate and refine the theory. This approach was useful for exploring the open-ended range of practices (Couldry 2012) and to more fully reconstruct the perspectives of the research subjects without applying pre-conceived concepts.

Spread over three rounds of data collection, I conducted five semi-structured interviews with members of mySociety, including its founder and former

CEO, two senior developers, a member of the international team, and a member of mySociety's research team. These interviews had an average length of one-and-a-half hours. A large part of each interview consisted in the reconstruction of a particular project. Each interviewee was asked to pick a project that illustrates her or his work and then walk me through the development process: What was the initial idea behind the project, what were the different steps and phases for implementing the project, what happened after the initial release? Reich (2013, 422) calls this 'reconstruction interviews' because it reconstructs "technological 'biographies'". Exploring the development of a project in-depth vividly illustrated the practices and routines of my interviewees. If not mentioned by the interviewees, specific questions about the role of data were asked. After one project was explored, I asked whether this was a 'typical' project and if there are very different examples. If there were, I explored those as well. Other questions addressed self-understandings (preferred job title, understanding of civic tech) and personal or organizational ambitions and values.

These interviews were complemented by 17 documents found online: mySociety's homepage (including the use of the Internet Wayback Machine to retrieve older versions), project specific websites, blog posts and forum discussions from Tony Bowden (who has been working for mySociety since 2009), other interviews given by different mySociety members to newspapers or bloggers (e.g. Townend 2008), as well as presentations given by members like Tom Steinberg or Dave Whiteland available online (UsNowFilm 2008; Arcopix 2014; IndigoTrust 2011; mySociety 2014). Some of these documents were included in the initial data sample, others were added later following the theoretical sampling. Moreover, I conducted ethnographic research on two separate conferences which were visited by several mySociety members: The

Open Knowledge Festival 2014 in Berlin and the Mozilla Festival 2015 in London. These conferences helped to get an impression of the larger community mySociety is involved with and provided a helpful guidance for both the interviews and the analysis.

How mySociety members imagine the affordances of structured data

In the following, I describe the imagined affordances members of mySociety hold around structured data. First, I will explore mySociety's mission and self-understanding in more detail to give a dense description of the broader ambitions and imaginaries that drive this organization. Then I show how data is used to facilitate this mission by describing four imagined affordances: deep linking into documents to engage citizens with the processes of governments; making the performance of governments legible to affect how they implement laws and public services; affecting its users' perceptions by demonstrating their impact to them; and scaling technological solutions to support a distributed form of agency.

mySociety's mission: Facilitating engagement

mySociety's self-proclaimed mission is to "help citizens demand better (…) our web tools and apps are breaking down the barriers around governments" (mySociety n.d.). Its tools are supposed to give "greater access for citizens to the work of government and the democratic process", which essentially means improving how publics can monitor and provide feedback to governments: "We believe that governments tend only to get better at serving the needs of citizens when citizens are capable of demanding better, creating a virtuous

circle that leads steadily to better government" (Cridge 2015). While this statement might imply advocacy for specific policy changes, mySociety understands its role as a 'mere' facilitator of civic engagement, not as an advocacy organization.

Members generally reject the idea of being a gatekeeper that stands between the citizens and their governments. Instead, they suggest to provide the means by which others are able to take actions more effectively: "What we do is present the facts: This is how your MP [Member of Parliament] voted, this is where the money went, this is what was said. It's then up to other people to do with that what they will, which might well be using it to promote a cause" (Interview: Research Team). The idealized and simplified scenario that members sketch out rhetorically to describe their role can be outlined as follows: Before mySociety enters the arena, citizens are apathetic and disengaged because engagement is too difficult and time-consuming due to high barriers raised by governments (in most cases unintentionally). mySociety identifies these barriers and then 'drops' its tools into the public arena to make engagement easier for citizens, which subsequently facilitates engagement between citizens and their governments and leads to better governance.

Underlying this approach is the assumption that *more means for citizens to provide feedback to elected representatives* leads to 'better' outcomes, i.e. *more representative* and therefore *more democratic* outcomes. Given its self-understanding as a 'facilitator', mySociety is not advocating for specific outcomes, but is concerned with the processes by which outcomes are generated: "We're in favour of a vibrant, healthy, lively democracy. That means a rude and obnoxious place. Although we don't want to do that ourselves, it's entirely appropriate that we should facilitate other people to" (Steinberg quoted in Krotoski 2010). This approach builds on the principles of open source culture

(Kelty 2008; Lewis 2012b). "The essence of open source," Weber (2004, 56) describes, "is not the software. It is the *process* by which software is created". As the ideas and practices of the 'open source process' were increasingly applied outside of software development (most prominently with Wikipedia) they formed the basis for a larger technological and cultural phenomenon that Jenkins (2006) has described as participatory culture, a culture "which posits that knowledge is richest and most accurate when it reflects the pooled inputs of a distributed population, as opposed to the expertise of a single agent" (Lewis 2012b, 847).

mySociety builds on previous forms of participatory culture and has particularly strong connections to technology-driven open data initiatives and rights-based open government or freedom of information initiatives, both of which are interested in applying the 'logic of open participation' (Lewis 2012b) to institutionalized politics (Schrock 2016; Janssen 2012; Yu and Robinson 2012). mySociety was an early supporter of open data in the UK and its founder was part of a group that articulated the '8 principles of open government data' (OpenGovData.org 2007). It also promoted freedom of information laws through its website WhatDoTheyKnow, which helps users to submit FOI requests to public institutions, and advocated for strong FOI legislations. Yet despite this strong connection, advocating for open data and FOI is not mySociety's main purpose, they are rather perceived as "resources that mySociety needed to function" (Interview: Former CEO). This is because mySociety does not just build on participatory culture, it also *extends* it in important ways. Participatory culture relied on the connectivity of internet technologies to establish new forms of governance based on collaboration and sharing. As Lewis (2012b, 848) describes, participatory culture is based on a forging of technology and culture, in that digitalization "*enables* greater user

participation on a seemingly infinite order, and the socio-cultural context of this technology has *encouraged* greater participation to achieve normative aims of collective wisdom and well-being".

mySociety similarly wants to create a more collaborative and participatory process for achieving better outcomes, but it does not solely rely on the connectivity enabled by internet technologies and ways of 'coordinating collaborations' (Kelty 2008). Its civic tech applications are not primarily about connecting people, but about *facilitating* them, i.e. enabling them to engage with governments in ways that go beyond 'mere' connectivity or access to information. In other words, mySociety is extending participatory culture by drawing attention to the conditions that would allow and encourage people to participate. Civic tech at mySociety is essentially about feasibility, in that it aims at making engagement more feasible for citizens by removing frictions such as needing to find out who represents them in parliament and how to contact them.

mySociety's imagined affordances of data are closely tied to this broader mission of creating a participatory culture. To illustrate the fundamental importance of structured data for mySociety, a member uses the analogy of cooking ingredients:

> If the useful thing is a cake that people want to eat, you're interested in the ingredients (…) But you don't want raw ingredients like wheat, you need the flour. Some processing has to be done to the ingredients before they are ingredients that you can sensibly make a cake with (…) Without the structured data, you wouldn't be able to offer that service and until you can offer that service you couldn't really prove that the demand for it would be so great. (Interview: International Team)

This suggests that structured data, if 'served' correctly, would increase engagement and subsequently alter the relationship between citizens and their

governments because it shows that the 'demand is great'. In the following, I unpack this statement and describe the imagined affordances of data underlying it.

Improving engagement with governments: Deep linking

One of the most popular services on mySociety's parliamentary monitoring website TheyWorkForYou are email alerts. For example, users can search for keywords in parliamentary discussions and then sign up to regularly receive emails informing them when their keyword comes up in future discussions. This service was significant because British parliaments used to publish transcripts of speeches as PDF files. To monitor keywords or what individual MPs are saying in parliamentary discussions, one needed to download these PDFs regularly and search through them individually. The email alerts at TheyWorkForYou turned this monitoring process into something people are able to do along the way, without investing considerable time and effort.

The key practice behind this service is the idea of 'deep linking': "The idea that there is councilor Jones who said 'This is what we have to do in my home town!' and you can cite it directly" (Interview: International Team) – similar to Twitter, where every individual Tweet has its own URL which can be shared or embedded on other websites. For deep linking, mySociety (n.d.) considers transcripts "made of nicely structured data (…) hard to beat". If documents are in a format that does not allow deep linking, "you can't cite, you can't share, you can't show specific utterances (…) that pretty much stops details in documents being called out in debates" (Whiteland in mySociety 2016). Today, it uses a data standard for modeling parliamentary speeches called Akoma Ntoso. It allows granular filtering (everything *this* person has said), an analysis

of the speakers' behavior (what was said, how and when?), and the ability to link speakers and what they say to other statements and events: "We would like to see people (…) build sites like 'all public statements by the Prime Minister'" (mySociety n.d.). This creates a level of "semantic understanding" with an "awareness of speakers" (Interview: International Team) that is necessary for the services mySociety develops.

The ability to link deeply into a document using structured data is considered "really important in public discourses about documents" (Interview: International Team). More fundamentally, documents are considered important because mySociety is "concerned with the process of government and most government (…) is actually the business of making laws, and laws traditionally have been written down" (Interview: International Team).

From the perspective of mySociety members, enabling deep linking into documents and providing services around it improves the public's awareness of, and engagement with, the businesses of governments (as those businesses are captured in documents). As mySociety explains:

> Transcripts are a kind of oil that greases the wheels of well-functioning societies. They let people discover when powerful people have made pronouncements that affect less powerful people. We believe that by making transcripts function better, more people will end up learning about decisions and opinions that affect their lives. (mySociety n.d.)

In the interviews, members also frequently referred to a claim by mySociety's founder: "everything you can and cannot do in your life has been decided by more powerful people in a meeting" (cf. Whiteland in mySociety 2016). Deep linking is about improving access to such meetings to help figure out "who was responsible for things" (Interview: Former CEO) by improving the

accessibility of those documents which capture what was said by whom. Put simply, deep linking should help to keep track of where and when powerful people make decisions.

Taken together, mySociety is imagining the affordances of structured data in this case as a basis for a *document-driven monitoring tool* to help engage the public with the businesses and decision-making processes of governments. Through deep linking, structured data is imagined to "even [affect] an apathetic population, it affects the way that they behave and what they know about what's going on in their own society" (Whiteland in Arcopix 2014). This makes turning documents into structured information a central part of mySociety's mission of creating a more participatory culture. Importantly, deep linking is more likely to facilitate individuals who are already highly engaged and interested in the processes of governments to begin with, either privately or professionally (e.g. journalists or activist groups). As I will discuss below, mySociety also imagines data to help engaging citizens who are not necessarily interested in politics or feel disengaged and powerless.

Mediating between governments and their publics

On FixMyStreet and WhatDoTheyKnow, mySociety collects all the problem reports or freedom of information requests by its users, tracks the responses by public institutions, and makes both publicly accessible. By making the resulting databases public, mySociety creates new forms of legibility and assessability. The database on FixMyStreet enables the analysis of a city's infrastructure problems by the public and makes the performance of governments, e.g. how fast they fix problems in specific regions, legible. WhatDoTheyKnow similarly allows an assessment of how different public

institutions respond to FOI requests and subsequently how the law is being practiced.

One reason for mySociety to 'aggressively' impose transparency on governments is its reliance on their cooperation. Its applications are built *on top* of services provided by governments and cannot exist independently without them. If public institutions refuse to cooperate, for example by ignoring reports sent via FixMyStreet, their tools would be of little or no use. By imposing transparency and allowing the public to assess their performance, mySociety makes it more difficult for authorities to ignore them. It is a way of pushing institutions to cooperate and to adopt mySociety's emphasis on user-friendliness and accessibility (see below). This is described as "the one bit of activism that we occasionally engage in" (Interview: Research Team).

A less obvious aspect of this 'one bit of activism' is how mySociety's use of data enables it to mediate between the bureaucratic and legal processes of governments and the users of its applications. The way mySociety's tools work is usually not a direct reflection of how public institutions or legislations are working:

> We think about the aim of the software as not being necessarily to model exactly the processes as they already exist in the world, but to make the software embody a slightly better way of doing things. (Interview: Senior Developer 2)

In this sense, mySociety's tools are "intended to provoke some friction" (Interview: Senior Developer 2). In the cases of FixMyStreet and WhatDoTheyKnow, the laws and regulations did not prohibit the publication of problem reports or FOI requests, but it also was not something that was suggested. In addition, mySociety reinterprets how services and laws should be implemented in smaller ways, for example by forwarding reports to councils

via email, while councils prefer people to use a form on their website; or by allowing users to send FOI requests to institutions that were not subject to the FOI law on WhatDoTheyKnow (mySociety 2011).

mySociety developer Tony Bowden (2010) described this principle as 'dreaming out loud': "simply act the way you want the world to be, then wait for reality to catch up" – by finding ways to make institutions cooperate and push 'reality' in that direction. This 'dreaming out loud' principle fundamentally depends on mySociety's use of data to make the performance of governments legible and assessable for the public.

However, mySociety cannot simply invent new ways of how governments should work and then impose this vision on institutions. While they are intended to create frictions, the software tools and data structures mySociety develops need to reflect the real-world processes they are supposed to represent to a large enough degree "to actually work" (Interview: Senior Developer 2). Instead of 'reinventing' government, mySociety's tools are compatible with *existing* services and legislations, but simultaneously utilize data to asses and affect how they are being *implemented*: "that gap between the way the law works, the way institutions implement the law and the way perhaps it might be ideal from a citizen's point of view is an important one and one where I think it's significant to build an artifact that demonstrates that" (Interview: Senior Developer 2). Using data to demonstrate and potentially close this gap is mySociety's definition of 'empowerment': people are empowered by giving them tools which enable them to "see and be able to do what they are legally entitled to as easily as possible" (Interview: Research Team).

This understanding of empowerment encapsulates how mySociety imagines to affect the relationship between citizens and their governments through data. It uses data to facilitate the use of preexisting rights and services, not to fundamentally change them. It envisions its tools to act like a 'layer' that translates the bureaucratic and legal procedures followed by public institutions into user-friendly interfaces with accessible language for citizens. They are supposed to both reflect existing processes *and* visions of how these processes should 'ideally' work. The legibility and assessability created by data is imagined by mySociety to affect the implementation of rights and public services in ways intended to make engagement easier for citizens and thus 'push reality' closer to its vision of a more participatory culture.

Changing perceptions: Providing a sense of agency

While facilitating already interested users via deep linking is important to mySociety (see above), it ultimately aims at engaging people who are usually not engaged. mySociety aims at a more long-term change in perception about "what is normal rather than what is exceptional" (Interview: Former CEO). Using TripAdvisor as an example, the former CEO explains that it "has caused a massive power shift in the hotel industry from the people who run the hotels towards people who stay in hotels", even though it is not a "campaign for better hotels". Services such as TripAdvisor would change people's expectations about things like going to hotels if they are popular enough, i.e. widely used (cf. Steinberg in IndigoTrust 2011). In a similar vein, mySociety wants its tools to be "popular rather than idealistic" (Interview: Former CEO) and has a strong emphasis on usability, i.e. on making sure its tools are easy to use and provide "the same level of service that the best

[commercial] websites we use nowadays would have" (Interview: Senior Developer 1).

In order to simultaneously achieve popularity *and* promote a more participatory culture, mySociety tries to 'harness self-interest' (Bowden 2014b) by generating "public good from private desire" (Steinberg in Nestoria 2008). It aims at developing services that address individual end-users, but use the data gathered from the engagement of these users to create "public value" (Steinberg in UsNowFilm 2008) on top of it. Crowdsourced databases such as those created on FixMyStreet or WhatDoTheyKnow are key to this. FixMyStreet helps individuals to fix their specific problem, but it also collects all the reported problems to create a public database about local problems that can be useful for others, for example local journalists who can sign up to get email alerts for problems reported in a specific region.

By bridging individuals with collectives in this way and by emphasizing ease-of-use, mySociety ultimately wants to affect the perception of users about their own agency. For example, when a problem reported via FixMyStreet has been fixed, it will be removed from the public city map and the user who reported it will receive an email letting her know and encouraging her to report more, or to try out mySociety's other projects for contacting her representatives. This way, mySociety is not only solving an individual's problem, but is also trying to demonstrate the public value created by its action. This is intended to give users a "sense of agency (…) some ability to change their environment" (Interview: Senior Developer 2) by demonstrating that their actions do have an impact:

> what we often see as apathy is really just learned helplessness. People feel powerless, because they don't believe they can make a difference. And the best way to change that is not to argue with them (…) It's to simply show them that they *do* actually have power. That what they do *can* have an effect, not only for themselves but for people around them. (Bowden 2014b)

Here, the legibility and assessability of crowdsourced data is imagined to create a more abstract, cultural and psychological change in perception.

Taken together, the crowdsourced databases mySociety creates are imagined to affect both governments and the public. On the one hand, they enable mySociety to push governments and influence how laws and public services are being implemented, as they make the performance of governments legible in new ways (see above). On the other hand, it uses the data generated by its users in ways intended to change how they feel about engagement by providing easy-to-use tools and by demonstrating that their actions had an impact. Databases are imagined to affect emotions and perceptions in order to advance a vision of a more participatory culture, which illustrates the importance of emotions and perceptions around data (cf. Kennedy and Hill 2017). As I show in the next section, data is also key to mySociety's ambition to support the development of similar applications elsewhere.

Scaling civic tech: Supporting a distributed form of agency

The bulk of mySociety's work today is based on collaborations with groups in other countries to create national versions of projects originally created in the UK, especially Alaveteli (an international version of its right-to-know site WhatDoTheyKnow) and tools for parliamentary monitoring. In part, this international orientation is driven by funding opportunities, but also because

numerous groups in other countries expressed a desire to have customized versions of mySociety's UK tools. While partly self-interested, members share the desire to support groups that similarly want to promote a form of participatory culture elsewhere. They think of themselves not merely as 'tool suppliers', but as part of an international civic tech community.

mySociety essentially wants to promote its values and practices in many different contexts by supporting local groups with developing customized versions of its tools. To achieve this, it has to accommodate the fact that "people in different places care about different aspects of politics. In some countries what really counts is how politicians vote, in others the crux is campaign finance contributions" (Steinberg 2012). Data is seen as both a key problem and a solution to this approach. One of the main obstacles for transferring an existing application to a new country or for building new civic tech applications from scratch is the design of consistent data models that adequately reflect the structures and legislations of the respective national government. A data model designed to capture the British parliamentary system cannot simply be transferred to another country. If they are not already available (as open data), developing such data models is complex.

Therefore, mySociety is one of the founding members of Poplus, a 'global civic tech federation' of organizations similar to mySociety, like Code for America in the US or g0v in Taiwan. The basic idea that drives Poplus is that "pretty much every tool in the civic and democratic space can be broken down into some parts that are universal, with usually only a little bit of local glue holding them together":

> although pretty much every Parliament has different processes, they're still dealing with *the same raw ingredients* – people, parties, speeches, bills, votes, etc. – and if we could create standardised ways of modelling each of those things, it would be a lot less work for people to combine these in the way that makes most sense for their own situation. (Bowden 2014a, emphasize added)

In connection to Poplus, mySociety supports data standards for modelling government structures and develops EveryPolitician, a project that literally collects data about every politician in the world. In this project, mySociety makes "data editorial decision[s]" (Interview: International Team) about how the 'basic elements' that can be found in every government are represented in a consistent manner across countries to ensure tools can be easily deployed internationally.

On a technical level, mySociety is trying to reduce problems of scale through data structures and tools that standardize how common 'ingredients' of government systems are formalized. However, members perceive this not merely as a technological detail, but as a way to facilitate mySociety's vision of a more participatory culture elsewhere. By enabling other groups with similar values and social imaginaries to create their own, local versions of civic tech applications that serve their particular needs, mySociety is essentially trying to support a *distributed form of agency*. By promoting data standards and reusable tools, it tries to create the conditions necessary for supporting its values and approaches in as many different contexts as possible.

Conclusion: The cultural and historical situatedness of affordances

This article examined the imagined affordances members of mySociety hold around data to gain a deeper understanding of how civic technologists rely on

data to meet their own ends. mySociety is trying to facilitate civic engagement and, by extension, create a more participatory culture. Taken together, it is imagining the affordances of structured data in ways that are supposed to expand the agency of publics towards governments: ways intended to enable citizens to better influence and interact with governments or other powerful institutions. Four imagined affordances have been identified.

First, members of mySociety use structured data to make 'deep linking' into documents possible, thereby aiming to increase the publics' awareness of, and engagement with, the processes of governments. Second, crowdsourced databases enable mySociety to mediate between governments and their publics. By collecting data to monitor the performance of governments, and by making this data publicly available, mySociety is trying to push governments to cooperate with its applications and thereby affect how existing laws and public services are being implemented in supposedly more user-friendly, i.e. 'citizen oriented' ways. Third, members of mySociety are trying to use data to affect the perceptions and self-understandings of their users. Crowdsourced databases are supposed to help connect the individual with the collective by demonstrating to individual users that their actions have an 'impact' and create a kind of 'public value'. Fourth, members of mySociety are scaling technological solutions to support a distributed form of agency that should enable groups in other countries, which similarly want to promote a more participatory culture, to develop customized versions of mySociety tools that serve their particular needs.

The way mySociety is using data to apply notions of participatory culture to politics suggests that we should understand the role of civic technologists in the public arena as *facilitators of engagement*. Facilitating means that data is used in ways that are supposed to enable others to take actions themselves.

mySociety does not advocate specific outcomes (such as particular policy changes), but is concerned with how the *processes* by which outcomes are generated are designed. It aims to enable users to easily engage with authorities in order to ensure that decision-making "reflects the pooled inputs of a distributed population" (Lewis 2012b, 847). In other words, mySociety wants to influence the conditions by which others participate in the public arena without directing the public discourse or influencing policy making. At the same time, it sees its role as complementing actors who *do* want specific outcomes, like advocacy groups within civil society, or professional journalists who emphasize gatekeeping.

By showing how members of mySociety understand and use data to facilitate civic engagement and realize a more participatory culture, this article draws attention to the fact that "the political and democratic possibilities of data" (Milan and Van der Velden 2016, 8) cannot be determined in an abstract way. What data affords to whom does not only depend on the technological properties of data, but is fundamentally social, and both culturally and historically situated. This has implications for studying data activism in general and civic tech in particular. Related to data activism, the findings presented here invite us to extend Milan and Van der Velden's (2016) conceptualization of data activism as creating novel epistemic cultures around datafication within civil society. The epistemic cultures and related social imaginaries promoted by mySociety are not novel: the practices and imaginaries described here build on participatory culture, which itself has roots predating computer culture and notions of open source software (Tkacz 2012). mySociety's practices and epistemologies did not appear out of nowhere and the epistemic cultures they create are not necessarily novel, but they develop imagined affordances around data to *implement* these ideas in new ways. To understand imagined affordances,

historical trajectories are just as important as the new elements added by data activists.

Studying imagined affordances does have a lot of potential for further developing and refining notions of data activism. First, imagined affordances provide a useful angle for moving beyond broad dichotomies of pro-active and re-active forms of data activism (Milan and Van der Velden 2016) and for studying how the epistemic cultures developed by data activists are connected to data in nuanced ways. Especially when combined with a focus on practices, it is a powerful tool for illustrating the "distributions of agency and organising forces" (Tkacz 2012, 404) activists set in motion. Second, the article shows how imagined affordances can be used as an integrative framework for studying how data activists affect the distribution of knowledge and power. I described how mySociety is trying to utilize the legibility and assessability created by data to change the perceptions of users about their own agency. Subsequent studies could examine how the practices and social imaginaries of data activists interrelate or clash with the self-understanding and perception of different groups in the public arena, and how different imagined affordances by various types of users emerge around civic tech applications.

Applied to civic tech, the research approach developed here calls for a more differentiated picture of the civic tech sector. Civic tech organizations and their funders are very much interested in a "coherent and clearly articulated vision and sense of shared identity for civic tech" (Donohue 2016). Yet while the diverse actors in the field of civic tech might align inasmuch as they are all interested in open data, reusability and a vague sense of improving 'civic life', their interests might eventually clash. For example, European non-profits like mySociety or the Open Knowledge Foundation Germany advocate for governments to *copy* civic tech applications. At mySociety, this even goes as far

as stating that most mySociety projects "shouldn't need to exist at all" (Interview: Senior Developer 1). Elsewhere, civic tech is closely aligned with start-up culture and attempts to create new industries. Wanting public institutions to copy civic tech applications in order to change the relationship between citizens and these institutions, or favoring the 'government as a platform' paradigm (O'Reilly 2010) and wanting public institutions to 'step back' and foster ecosystems of for-profit services, has very different implications for civic life (Baack in DataDrivenJournalism.net 2016).

Pointing out such differences is also important for developing a more differentiated critique of civic tech. Civic tech and open data initiatives have been said to uncritically support a neoliberal agenda driven by commercial interests rather than government accountability or citizen empowerment (cf. Bates 2012; Slee 2012; Gregg 2015). Moreover, their focus on technological solutions would merely reinforce existing power relationships by 'empowering the empowered' (Gurstein 2011). While this critique is important, it does not equally apply to every actor in the civic tech sector. We should not discard the agency and intentions of non-profit organizations such as mySociety, who are aware of these discussions and conduct research to better understand the impact of their applications. Members of mySociety constantly experiment and explore ways to support and encourage civic engagement. In this sense, they are "potential 'laboratories'" (Hepp 2016, 929) that we should not ignore if we want to formulate normative principles for making processes of datafication more democratic and for creating more self-aware and agentic publics (Kennedy and Moss 2015).

The imagined affordances identified and described here can help to map differences among actors in the civic tech sector: Are other civic tech organizations relying on data in the same way, and do they promote similar

ends with them? Are actors outside the field of civic tech adopting them, and in what way? Which actor in the civic tech sector influences what field? Given civic tech's status as a 'pioneer community' (Hepp 2016), tackling these questions is one way to conduct empirical research to trace and understand how forms of civic engagement and activism are changing due to their growing reliance on data.

6. Practically engaged

The entanglements between data journalism and civic tech

Abstract: This article explores the entanglements between data journalists and civic technologists. Following an approach inspired by practice theory, it describes how they form a community that comes together through interlocking practices and complementary values and ambitions. Data journalists and civic technologists interlock along a continuum that oscillates between practices of facilitating (enabling others to take action themselves) and gatekeeping (being impactful and steer public debates). Depending on how much emphasis is put on either facilitating or gatekeeping, four different groups are identified that differ in how they position their work, in their professional self-understanding and in how they use data: Normalizers, Experimenters, Translators and Facilitators. The article concludes by suggesting that actors populating this community of practice can be described as flexible data professionals who aspire to work in a public interest. The findings illustrate how the progressive datafication of social life creates new entanglements between the field of journalism and civil society and we should pay more attention to such entanglements and the implications for increasingly datafied publics.

Introduction

In July 2009, Alan Rusbridger, editor of The Guardian at that time, gave a keynote on 'Why Journalism Matters' (in MediaStandardsTrust 2009). He opened with stating that newspapers used to be the gatekeepers of data and official information, but that has changed. He gave three examples: the problem-reporting website FixMyStreet, the parliamentary monitoring website TheyWorkForYou, and the hyperlocal website EveryBlock. For him, they are examples of changes in "how information is organized, personalized, ordered, stored, searched for, published, and shared". They are "dealing with facts, with statistics, with information about public life, politics and services" and have

many things in common with "conventional journalism". With a mix of excitement and uncertainty, he concluded: "I don't know if that is journalism or not. I don't know if that matters".

Today, the examples given by Rusbridger are commonly referred to as 'civic technologies': technologies developed by non-profits, companies or governments themselves trying to make it easier for citizens to engage with their governments. Rooted in small volunteer experiments in the early 2000s in the US and the UK, the civic tech sector has grown substantially and one can find civic tech organizations in almost every part of the world today: non-profits funded by foundations like the Open Society Foundation, startup companies, initiatives by huge corporations like Google or Microsoft, and by governments (especially in the US).

Rusbridger's speech is one of many examples to show that civic tech had a close relationship with journalism early on: journalists have increasingly adopted practices and ideas from civic tech, use civic tech applications for their own investigations, and occasionally directly cooperate with civic technologists; while civic tech organizations like mySociety, the non-profit organization that developed FixMyStreet and TheyWorkForYou, encouraged journalists to use its tools in various ways or sought cooperation with media organizations. For Alexander Howard (2014c, 64), who has been active in both fields, the parallels "to what civic hackers are doing and what data journalists are working on is inescapable". Both are related to the 'civic' and public life, advocated for freedom of information and transparency, trying to provide a public service that empowers citizens, and both fundamentally rely on data (Schrock 2016).

Yet we know little about how civic technologists and data journalists relate to each other's work, how they complement each other, and where they differ. To a large extent, research in journalism studies has concentrated on how data journalism is practiced *within* newsrooms (Fink and Anderson 2015). How data journalism is shaped by, and shaping, other fields in the technology sector was a relatively marginal concern, despite the fact that it is largely acknowledged that the news-making-process is "increasingly shaped by networked forces (…) that span multiple professional identities, information ideologies, and assumptions about how news and public life intersect" (Ananny and Crawford 2015, 192–93). It includes not just journalists, but a great "diversity of actors, discourses and relationships" (Domingo, Masip, and Costera Meijer 2015, 53) that influence how news is found, produced and circulated. Research on civic tech, on the other hand, concentrated on tracing its historical roots (Schrock 2016), understanding the users of civic tech applications (Cantijoch, Galandini, and Gibson 2016) or exploring its relationship to neoliberal government agendas (Bates 2012). Its relationship with data journalism has not been a primary focus.

Studying how data journalism and civic tech complement each other is relevant not just because there is a direct cooperation and overlap between them, but because this relationship is shaping how journalism and forms of civic engagement are responding to the progressive datafication of social life (van Dijck 2014). As a powerful emerging knowledge logic, datafication fundamentally affects how we collectively make sense of, and engage in our social worlds. Both data journalists and civic technologists aim to produce knowledge in the public interest and their entanglements affect the wider process of knowledge production and circulation in datafied publics. How

does not just the reliance on data, but the way both fields *complement* each other, affect how they work?

Instead of looking at professional journalists and civic technologists separately as distinct fields, this article therefore looks beyond organizational boundaries and brings together perspectives from both equally to study the "open-ended range of practices" (Couldry 2004, 117) across these fields. It presents findings from a qualitative case study which explored how data journalists and civic technologists are making sense of their own practice and how they exchange, pick up, modify, reject and ultimately relate to each other. First, I conceptualize these groups as forming a community of practice or figuration that revolves around data-related practices and overlapping social imaginaries. Then, four interconnected groups within this figuration are presented: Normalizers, Experimenters, Translators and Facilitators.

Methods: A focus on overlapping and diverging practices

To study the entanglements between data journalists and civic technologists, I took inspiration from practice-focused research paradigms that try to avoid "any apriorisms about the roles and practices of the multiplicity of actors" (Domingo, Masip, and Costera Meijer 2015, 54). A focus on practices has been suggested by a number of researchers to avoid *a priori* delineation of actors based on predefined categories. Instead, it encourages researchers to be open to the full range of "what people are doing and how they categorize what they are doing" without predefining their actions in categories like 'consumption' "whether or not that is how actors see their actions" (Couldry 2004, 125).

The findings presented here are the culmination of three distinct case studies on civic technologists and data journalists that mutually informed each other. The first case study explored the social imaginaries of the Open Knowledge Foundation Germany (OKF DE, Baack 2015a), a non-profit organization and the most influential actor in the German open data movement. The second case study explored how civic technologists at the British NGO mySociety use data to realize their goals (Baack, 2018). The focus on mySociety was informed by the first case study: mySociety is one of the oldest and internationally most influential civic tech organizations today that also had a huge impact on the German civic tech scene. Even some data journalists interviewed for this study pointed out its influence: "Back in 2008 and 2009, they [mySociety] absolutely were inspiring role models for me and other people I worked with" (Interview April 20 2016). mySociety represents an international 'best practice' and exemplifies some of the main characteristics of civic tech.

The third case study, whose findings are presented in this article, compared the practices and perspectives of civic technologists with those of data journalists in Germany, mostly in Berlin. The data journalism and civic tech scene in Berlin was particularly interesting, first, because there is a high concentration of influential German and/or European data journalists (e.g. from Zeit Online or Journalism++) and a very active civic tech scene (the OKF DE is stationed in Berlin). Second, the exchange and collaboration between local data journalists and civic technologists is frequent and continuous, in part thanks to numerous local workshops or events like Hacks/Hackers (Lewis and Usher 2014). Third, the local data journalism and civic tech communities in Berlin are well connected within transnational communities, both online and through international events like the Data Harvest conference. The influence of the local scene and their embeddedness

in transnational networks suggest that the findings presented here should be applicable and generalizable beyond the local settings.

I employed a qualitative mixed method approach. First, 29 interviews with 27 interviewees were conducted, the majority (22) face-to-face in Germany or in the Netherlands, the rest via Skype. I interviewed members of civic tech organizations, data journalists working in different organizational settings in Germany, and a couple of actors active in both fields simultaneously. Key were questions about the professional identity and an in-depth exploration of a particular project. Second, most interviews included participatory mapping (Emmel 2008). The interviewees were given a blank piece of paper (or a link to a web application) and asked to draw a mind map of all the communities they belong to and other groups that influence their work. During the process, questions were asked to explore how they relate to the different groups they mention and how they see their role in relation to them. This helped to reconstruct their subjective sense of belonging and processes of 'communitization' across organizational boundaries (Hepp, Berg, and Roitsch 2014). Third, I collected a range of online materials about each interviewee, such as social media profiles, interviews they gave elsewhere, articles they wrote or were written about them, slides and videos of presentations they gave. This material was collected and categorized to prepare each interview, and some of it was included in the analysis. Fourth, I conducted ethnographic research on several conferences (e.g. the Mozilla Festival 2015) and numerous local workshops in Berlin.

The analysis followed a constructivist grounded theory approach (Charmaz 2006) and was conducted with the help of the TAMS Analyzer (Weinstein 2006), a tool for qualitative data analysis. Grounded theory was chosen because its principle of theoretical sampling helped to explore the open-ended

range of practices without applying pre-conceived categories. Theoretical sampling means that an initial data sample was continuously expanded with new data to systematically elaborate and refine the theory. *Constructivist* grounded theory in particular has a focus on capturing actions and fluid processes rather than static concepts, which was also the main focus of this study.

A community of interlocking practices: Three examples

When we look across the field of data journalism and civic tech, the connections between individuals rarely revolve around sustained and institutionalized engagement. The exchanges are mostly informal and shaped by what Shove (2003) would describe as an order and form of collaboration determined by the fit of one practice with another (cf. Couldry 2012, 42–43). Groups working in different organizational settings are able complement each other to some extent, allowing a seemingly seamless transition between civic tech and data journalistic projects. Let me illustrate this with three concrete examples.

YourNextMP: Providing infrastructures for journalists

For the elections of the British parliament in 2015, the British civic tech organization mySociety built YourNextMP, a crowdsourcing platform to collect details about every candidate in the upcoming election. At the time, there was no central database provided by the government and only a couple of commercial databases offered this information. The goal was to create an open database that everybody could use without restrictions. This would make it easier for others to "interrogate" candidates: "say you wanted to survey your

candidates, or analyze their use of social media, or just find out who it is that's standing in your area, then that's very difficult to do unless there is a central database which does include all the candidates and all the constituencies" (Interview May 15 2015).

To further support such 'interrogations', mySociety linked each candidate to information provided by its parliamentary monitoring website TheyWorkForYou, which gives detailed information about each candidate's voting behavior and access to their speeches in parliament. Moreover, it provided widgets that were used by some local journalists and activist groups to embed a list of candidates in their constituency on their websites. The database was widely used during the elections. mySociety essentially provided an infrastructure consisting of a database, tools for investigating (TheyWorkForYou) and presenting information (widgets) for journalists and activist groups.

Code for Germany: Local laboratories of informal collaboration and learning

In February 2014, the OKF DE launched the 'Code for Germany' initiative as part of the 'Code for All' network, which fosters civic tech ecosystems on city levels. To date, 'Open Knowledge Labs' (OK Labs) exist in more than 20 cities across Germany. These Labs became one of the most important spaces for civic technologists and data journalists to meet and cooperate. Different to similar events like Hacks/Hackers (Lewis and Usher 2014), OK Labs are more regular and participants work on joint projects over longer periods of time. This creates a "knowledge infrastructure (…) which allows local match making (…) a meeting spot where different professions come together" (Interview March 11 2016).

As an example, the local newspaper Heilbronner Stimme and the local OK lab managed to establish a continuous, sustained collaboration. Members of the OK lab developed the technology, while journalists helped organizing, directing and publishing projects: "With my journalistic education, I was able to provide them with some structure and direction" (Interview May 13 2016). An example is a tool that allows users to specify their location to get detailed information about the quality of their local tap water. Civic technologists and journalists developed the tool together and the final product appeared on the newspaper's website. The journalists profited from the technological expertise, ideas and largely free labor of civic technologists; while the technologists were given a space to meet and feedback from journalists about the relevance and usability of their ideas: "It was a good combination (…) people from the open data community developed tools in their free time and we tried to get an audience for them and payed the developers sometimes" (Interview May 13 2016).

MinorInterpellations.de: Turning an investigation into an app

In Germany, members of parliament can pose questions to federal or local governments in so called 'minor interpellations'. The German Green party used this mechanism to get data about the condition of every railroad bridge in Germany. However, the geolocations were encoded in a format developed by the German railway company Deutsche Bahn. Journalists at Zeit Online took the data and translated this format back to standard latitude-longitude coordinates. After using this data for a story, the journalists were keen to share it: "We actually freed this data (…) It was really important for us to also publish it, so that other people can work with it as well" (Interview May 25 2016).

Later, this investigation was one of the inspirations for the OKF DE to develop KleineAnfragen.de (~'MinorInterpellations.de'), a portal that collects data provided by minor interpellations in federal and local parliaments. It allows users to search through minor interpellations across all German parliaments and offers email alerts for keywords. KleineAnfragen.de quickly became a standard tool for many of the journalists interviewed for this study.

Related to this example is how journalists and civic technologists complement each other's information rights. The interviews provided several examples where journalists used their exclusive information rights to acquire data to then share it with civic technologists in joint projects. Meanwhile, the advocacy of civic technologists for stronger freedom of information legislations is much appreciated by journalists as they are "unlocking opportunities" (Interview May 26 2016) for them: "they are engaging in fights we would have to engage in otherwise" (Interview May 18 2016).

A figuration based on complementary skills, attitudes and ambitions

The examples above help to illustrate how specific projects (like YourNextMP), shared spaces and institutional support (Code for Germany), as well as engaged individuals help to bridge different organizational backgrounds and allow data journalists and civic technologists to complement each other. Underlying these examples are three broader themes that characterize the communities they form:

1. *Overlapping and transferable skills*: Technological skills like writing scrapers, cleaning, analyzing and visualizing data are essential for both groups, and they rely on the same tools and programming libraries for their work.

The example of KleineAnfragen.de further illustrates an overlap in the use of information rights: knowing what can be asked (by law), what kind of data can be requested and where is essential for both data journalists and civic technologists. As an interviewee describes, the consequence is that "you really have people who come together in [Code for Germany] meetings and realize 'Hey, we are actually doing the same thing'" (Interview March 11 2016).

2. *Commitment to learning and open source culture*: Because of the overlap in skills, data journalists and civic technologists primarily form learning communities that come together to learn new or refined ways to work with data. They share a strong motivation to 'get better' and enjoy tinkering and exploring technology. They generally describe their work as a continuous learning experience: "What they [data journalists and civic technologists] share is a love for data and that they are not afraid to learn something new" (Interview February 18 2016). Directly related to this is a joint commitment to open source culture (Kelty 2008). As Coddington (2015, 333) noted, the principles of open source "have been an important common ground for bringing together 'hacks' (journalists) and 'hackers' (technologists)". They enjoy sharing experiences (e.g. with free tutorials, 'behind the scenes' articles, workshops), and they are committed to transparency by sharing source code and if possible their data (as in the example of the train bridges investigation). Both data journalists and civic technologists emphasized the collaborative and friendly nature among their peers (across institutions).

3. *Complementary ambitions*: The examples above illustrate how both groups easily conceive their work as complementary. KleineAnfragen.de shows

> how investigations by journalists can spark ideas for civic tech applications; YourNextMP shows how civic technologists understand their work as providing support for journalists or other activist groups. The collaboration at the Heilbronner Stimme can be described as a continuous exchange of ideas and a desire to complement each other. Underlying this ability to conceive each other as complementary is a shared 'sense of moral order' (Taylor 2004). Both data journalists and civic technologists understand their work as a public service that holds powerful people or institutions accountable and supports an active and informed citizenry: "We have similar ambitions (…) it is about the empowerment of users, about giving them tools to search through and use the data themselves rather than just giving them a story or – from the perspective of activists – whatever authorities give them" (Interview September 23 2016).

The resulting community can be understood as a community of practice whose members are "informally bound by what they do together (…) and by what they have learned through their mutual engagement in these activities" (Wenger 1998). More precisely, the practices of individuals in this community 'intermesh' and therefore form a *figuration* (Couldry and Hepp 2017). Figurations are connections between human actors based on meaning and interlocking practices. They are "*relations of interdependence*" that describe the "complex ways of interweaving human beings" (Couldry and Hepp 2017, 59). The boundaries of a figuration are defined by the shared meaning that the individuals involved produce "through their interrelated social practices, which is also the basis of their mutual orientation to each other" (Couldry and Hepp 2017, 63). Actors within a figuration are 'intermeshing' in the sense that their paths "are comprised of interlocking *practices* (…) that interlock because,

as meanings, they are in a mutual relationship, answering, inviting, challenging, questioning and so on" (Couldry and Hepp 2017, 65).

Shared practices, different identities: From facilitating to gatekeeping

Figurations are never 'flat', they are characterized by a distinctive constellation of actors with an orientation towards a shared purpose, or what Couldry and Hepp (2017, 66) call a 'relevance frame'. The figurations formed by data journalists and civic technologists involve complex distinctions that go beyond a simple contrast between 'civic tech' and 'data journalism'. At closer inspection, we can identify four different groups within this figuration. Each group represents a distinct articulation of a shared repertory of "images, stories, and actions" (Mansell 2012, 33) that that comprises practices of *facilitating* and *gatekeeping*:

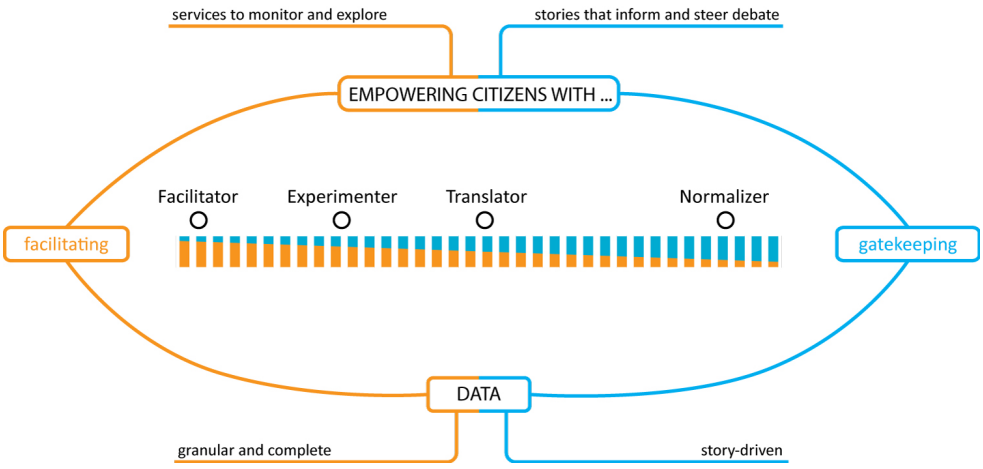


Figure 2: Figuration formed by data journalists and civic technologists.

Facilitating means to provide users with services that help them to take action themselves: exploring how an issue affects their personal situation, offering 'decision-making tools' (Parasie and Dagiral 2013, 864), engaging them with their governments. Put to the extreme, actors on this end of the spectrum want to avoid standing between the citizens and their governments and merely facilitate their exchange. Practices of facilitating are rooted in open source or participatory culture (Jenkins 2006) and as such follow the logic of open participation "of achieving a more engaged, representative, and collectively intelligent society" (Lewis 2012b, 848).

Gatekeeping, on the other hand, refers to the traditional journalistic role-model of being a gatekeeper for publicly relevant information. Actors on this end of the spectrum emphasize storytelling and impact. They want to inform their audiences about important events, provide guidance for public debate and affect policy making by amplifying public issues or misconduct. Their professional legitimacy builds on actively standing between the public and their governments (as the 'Fourth Estate').

Depending on how strongly they lean towards facilitating or gatekeeping, each group can be delineated along three categories. First, they differ in how they *position* themselves professionally. Towards gatekeeping, actors position themselves closer to the autonomous pole of professional journalism (Benson and Neveu 2005). With a stronger emphasis on facilitating, the lines between journalism and the wider technology sector become increasingly blurry. Second, they differ in their *self-understanding*: some understand their work as investigative journalism, others negotiate between journalism and civic tech and so forth. Third, they differ in their *data usage*. The distinctions here overlap with the different epistemologies among data journalists identified by Parasie and Dagiral (2013). Actors interested primarily in facilitating "put the emphasis

on data granularity and completeness" (Parasie and Dagiral 2013, 863). Data ought to be 'breakdown-able' so it can be analyzed and explored by others. The use of data by gatekeepers is story-driven: They use data to prove or falsify a hypothesis or to show the scope of a public issue to "inform public debate and influence the political agenda" (Parasie and Dagiral 2013, 860).

Note that the groups represent ideal types: abstractions and generalizations which are useful to categorize individuals without necessarily reflecting their own identity. In reality, the boundaries between these groups are blurry, but every interviewee had a clear bias that the classificatory scheme presented here captures. The scheme was developed by identifying the organizational and educational backgrounds of individuals as the most significant factor for determining similarities and differences among them. For example, journalists with a formal education in journalism working in established news media clearly position themselves towards gatekeeping, while those who emphasize facilitating have stronger roots in the technology sector and tend to work in more non-traditional setups like startups. Following grounded theory's principle of theoretical sampling, the classifications were developed by systematically exploring how data journalists in different organizational settings and with different educational backgrounds overlap or differ. The terminology was developed by the author on the basis of this analysis to capture how individuals understand and position their work. In the following, I will describe each group in detail.

Table 1: Different groups within the figuration.

Group	Positioning	Self-understanding	Data usage	Working environments
Normalizers	Position themselves close to the autonomous pole of professional journalism and engage in boundary work	Traditional investigative journalists	Support investigative reporting	Established national news media organizations
Experimenters	Work within journalism but strong positioning in technology sector, question boundaries	'Technologists' working in journalism, problematize and renegotiate journalism	Create personalized services that help to make decisions or form the basis for public debates	Non-traditional and often experimental working environments
Translators	Well connected to both journalism and civic tech, stress complementarities and similarities	Simultaneously as journalists and technology activists	Combine investigative journalism with civic tech, linking gatekeeping and facilitating	Non-traditional working environments at the intersection between journalism and civic tech
Facilitators	Position themselves in larger technology and non-profit sector	Enabler of civic engagement and gatekeeping, tool-supplier for other organizations	Facilitate engagement and gatekeeping to promote a form of participatory culture	Non-profit organizations

Normalizers

> I like data journalism because it's a return to the old virtues of journalism (…) I basically found everything I like about journalism in data journalism. That's why I got excited about it. (Interview September 23 2016)

Normalizers work in established national news media and emphasize continuity. They position themselves close to the autonomous pole of journalism and reflect a self-understanding that builds on a firm distinction between the 'inside' and the 'outside' of professional journalism. They think

of data journalism primarily as a "methodological competence" (Interview May 26 2016) that helps to "supplement, routinize, or algorithmically expand the scope" (Anderson 2013b, 1008) of existing journalistic practices and routines: "For me, it's a method I learn and work with, like a toolkit I use to tell the best story possible" (Interview August 05 2016). While some do appreciate the label 'data journalist' to signal their methodological expertise, others reject it: "You wouldn't call a journalist doing lots of interviews an interview journalist (…) I'm just a journalist, without a prefix" (Interview August 05 2016). Several members of this group made this comparison, suggesting that data journalism is a method just as natural and 'normal' to journalism as interviews.

Normalizers strongly identify with notions of watchdog journalism and firmly position themselves within the critical-monitorial tradition that shapes journalists' professional identity in most Western countries: "the ideal of journalism acting as 'Fourth Estate', with journalists voicing criticism and holding powers to account and, in so doing, creating a critically minded citizenry" (Hanitzsch and Vos 2018, 154). As data journalists, they are interested in gathering and analyzing data to scrutinize the performance of governments or other powerful actors. They subsequently are not 'passive' monitors who only take action once they become aware of issues or misconduct, they "proactively scrutinize political and business leaders; they provide an independent critique of society and its institutions" (Hanitzsch and Vos 2018, 154).

An example is a story by Zeit Online on anti-refugee violence in Germany.[24] Journalists started with the hypothesis that authorities largely failed to identify and convict perpetrators. To test this hypothesis, journalists collected their own data and gathered information about the status and success of every investigation. During this process, they continuously checked if their hypothesis is confirmed, ready to cancel or readjust the investigation if it was falsified. The resulting story is a classic piece of investigative journalism consisting largely of text and a few visualizations that demonstrated the failure of authorities.

This story illustrates many characteristics typical to the work of Normalizers. First, it is proactive and investigative and uses data to uncover and scrutinize patterns to prove or falsify a specific hypothesis. Second, the use of data is story driven. Data is collected and structured in a way that helps answering the hypothesis and with the later visualization in mind. Third, the project discloses a previously unknown fact and uses data to both strengthen its truth-claim and to show the scope of the issue. As it is typical for investigative journalism, this work aims to be impactful by producing a 'moral outrage' (Ettema and Glasser 1998) of the public which forces authorities to react.

When it comes to civic technologists, Normalizers are quick to engage in boundary work and emphasize their professional expertise that sets them apart (like objectivity and impartiality, ability to identify relevance etc.). While acknowledging overlapping goals, they are careful to point out that they do not embrace the political activism of civic technologists and only cooperate

[24] http://www.zeit.de/politik/deutschland/2015-11/rechtsextremismus-fluechtlingsunterkuenfte-gewalt-gegen-fluechtlinge-justiz-taeter-urteile.

under the condition that they maintain control over what is published. While emphasizing such differences to delineate their work, Normalizers are not trying to devalue the work of civic technologists. It is common among Normalizers to monitor what is happening in the civic tech scene (mainly via Twitter) and to be in touch with local civic tech groups through OK Labs or other informal meetings like Hacks/Hackers. They are interested in exploring how the work of civic technologists can complement their own work as "equal partners" (Interview May 25 2016), but with clearly defined roles and outcomes.

Experimenters

> I think Adrian Holovaty said he tries to provide information that helps people make sense of their surroundings. That's what we are trying to do and if it's journalism or not I don't care. (Interview March 29 2016)

Put simply, Experimenters are 'technologists' who neither clearly identify with 'journalism' (in the sense that Normalizers do) nor civic tech, but are interested in expanding data-driven computational techniques in journalism. 'Journalist' is the official job title for most of them, but they are somewhat indifferent about terminology and problematize the meaning of journalism. While Normalizers think of data journalism as a method that helps to improve 'traditional' journalistic storytelling, Experimenters emphasize the technological and experimental dimension and understand it primarily as 'doing journalism with structured data'.

Most members of this group have roots in the technology sector and did not have a formal education in journalism, and even those who do have a strong affinity towards technology and are self-thought programmers. They work in 'non-traditional' setups like startups or highly flexible and independent

developer teams within media organizations – conditions which are in many cases similar to those of civic technologists. Their networks usually display a broad range of different actors: from journalism to different NGOs, foundations, and technology meetups. Journalism does play an important role for all of them, but for some it is just one area among many.

Experimenters de-emphasize gatekeeping in favor of personalized services that facilitate their audiences, providing them with a new access and legibility of certain phenomena:

> I don't care how people interpret it, but I want to present the facts to them. The only impact I strive for is to enable people to see something that was invisible to them (…) one of the most important and central insights of data journalism is this: we don't publish stories, we publish the tools we developed to understand the data ourselves. (Interview February 18 2016)

An essential aspect to this is personalization, i.e. the ambition to illustrate how a phenomenon affects the individual user so that "everyone can find the most relevant aspect for him- or herself" (Interview August 02 2016). As a result, Experimenters often rely on interactive maps, allowing users to zoom in or directly enter their post code. An example is the 'noise map' developed by the local newspaper Berliner Morgenpost.[25] Using data provided by the city, this interactive map of Berlin uses colors to illustrate different noise levels on a street-level. Users can enter their address to check the situation in front of their own house.

[25] http://interaktiv.morgenpost.de/laermkarte-berlin/.

By providing this type of individualized legibility and assessability to their audiences, Experimenters want to help them "making decisions" and provide a "basis of discussion" (Interview August 02 2016). Ultimately, Experimenters envision a "completely data-driven newspaper" (Interview April 20 2016), an idea first articulated by Adrian Holovaty (2006). It suggests that news media should become "trusted data hubs" (Lorenz, Kayser-Bril, and McGhee 2011) by using automation on a large scale to continuously collect data and provide a broad range of services giving "recommendations or predictions" (Interview April 20 2016). The prototypical example of this is EveryBlock in the US, originally created by Holovaty himself (Parasie and Dagiral 2013). Implicitly or explicitly, Experimenters share a belief that this type of personalized services will encourage citizens to engage in public issues because it shows how a complex and potentially abstract issue affects them personally.

While Normalizers use data from a storytelling perspective, Experimenters are interested in granularity and completeness that makes data 'breakdown-able' to the individual. As Parasie and Dagiral (2013, 863) noted in relation to 'programmer-journalists' in the US, they "believe that intelligibility is the result of affording access (…) to complete and granular data from which citizens are usually kept away". What matters most is the depth and scope of the legibility and assessability afforded by structured data rather than its ability to strengthen truth-claims (Normalizers).

Experimenters do not see their emphasis on facilitating in opposition to traditional journalistic gatekeeping. On the contrary, it is seen as an expansion that will strengthen it. This is especially true for projects where journalists collect data themselves to make a previously neglected issue visible and challenge official data collection routines or the lack thereof (cf. Gray,

Lämmerhirt, and Bounegru 2016). An example is the Migrants Files[26], a project that collected data on migrants who died on their way to Europe. It started with taking and structuring data that was already available on diverse sources and eventually collected new data by monitoring news articles and transferring relevant information into a database. The project was Pan-European: different newspapers across Europe used the data to create stories interesting for their national audiences. The data collection process was both driven by the desire to capture the scope of the issue and the possibilities for generating stories. In such instances, the data-driven approach of Experimenters complemented the gatekeeping focus of Normalizers. This way, the work of Experimenters overlaps with those of Translators (see below).

Experimenters see their work as much more overlapping and complementary to civic tech than Normalizers do. By de-emphasizing the storytelling-aspect of journalism in favor of facilitating services, it is at times difficult for them to tell the difference to civic tech:

> The question is: Can you call yourself a civic hacker when you work as a developer in journalism? I wouldn't call myself a civic hacker, but I also wouldn't mind if someone does. It's not all that wrong I think. (Interview March 03 2016)

The boundaries between civic tech and data journalism are blurrier for Experimenters. They work in the field of journalism, but their practices and self-understandings are much closer to civic tech. This causes some tensions: Some Normalizers are concerned about the de-emphasis on gatekeeping, while some Experimenters felt pressured to give their work a more 'journalistic'

[26] http://www.themigrantsfiles.com/.

outlook. However, there are numerous examples of how Experimenters are able to complement both Normalizers and civic technologists (as in the Migrants Files).

Translators

> I saw that journalism is the logical consequence of this activism if you want to reach a larger audience. (Interview April 22 2016)

Translators are individuals actively involved in both civic tech and data journalism. They are similar to Normalizers in that they are strongly committed to investigative watchdog journalism and recognize journalism and civic tech as distinct fields. However, rather than engaging in boundary work and emphasizing distance and impartiality, their engagement in journalism and civic tech is inseparable. They either come from the civic tech scene and experiment with extending investigative (data) journalism with civic tech's emphasis on facilitating, or they are journalists who see their involvement in civic tech as a useful extension of their work. Put simply, they seek ways to integrate civic tech in journalism and help to connect both groups wherever possible. They tend to position themselves in the 'hacker journalism' scene, or more generally at the intersection between technology and journalism.

Importantly, they do not work in traditional newsrooms but in settings that can be described as 'intersections' between both journalism and civic tech. To illustrate this, I will focus on Correctiv, a nonprofit investigative newsroom similar to ProPublica in the US. Like ProPublica, Correctiv wants to support investigative journalism by specializing on long-term investigations which are shared with other media outlets. However, it extends the ProPublica model with an educational agenda:

> We train people: we want to pass on our methods of investigation and help citizens access the information to which they are legally entitled. Our goal is to help citizens make society more transparent and so to foster democratic engagement. (CORRECTIV n.d.)

Correctiv promotes the idea that "everybody can be a journalist. What matters is the use of the right journalistic methods" (Jonathan Sachse in CORRECTIV 2015). Journalism is implicitly understood as a method everybody can learn. Civic tech applications blend nicely into that rationale as most of them are "research tools" (Interview April 22 2016) that help both journalists and citizens to investigate or utilize data.

Given a mission statement and understanding of journalism that can be read as an attempt to balance gatekeeping and facilitating, it is unsurprising that Correctiv has very close connections to both journalism and the civic tech sector. Its offices are located in the same building as the OKF DE and Correctiv hired two of its former members, which have both been interviewed for this study. They describe their involvement in journalism as a "logical consequence" (Interview April 22 2016). It allows them to reach a larger audience, but more importantly, it is seen as a way to put their advocacy into practice. For them, data journalism at Correctiv is "applied civic tech" (Interview March 30 2016): "First comes the goal of informing people and disclosing misconduct, but of course I can still use the same means as before and show people that it is good that those are available" (Interview April 22 2016).

Correctiv's attempt to link strengthening investigative journalism with educating citizens leads to an incorporation of practices and applications inspired by civic tech. Several investigations follow an approach developed by ProPublica: collecting data to shed light on a phenomenon that lacks

transparency, and then publish the database with tools to investigate it alongside the stories. An example is a large-scale investigation of nursing homes in Germany.[27] During this investigation, data was collected to get an overview of the sector and do an exploratory data analysis: How does the system work, what problems exist, what causes them and how could they be solved? The result is a series of stories, a TV documentary and an application that allows users to explore the data themselves. They can search for their city to get a map which highlights local nursing homes with some information, such as ratings or prices. Moreover, it offers a button to FragDenStaat.de (~AskTheState), the freedom of information website run by the OKF DE. Clicking on that button opens a pre-filled freedom of information request asking for nursing home reports. Once a request was successful, the new documents are automatically uploaded to the Correctiv's database.

This research tool is intended to help local news media to do their own local investigations, but it also clearly incorporates civic tech's emphasis on facilitating citizens and follows Correctiv's stance that everybody can be a journalist: "Being able to request more specific information, that's empowerment. It's exactly what mySociety or the Open Knowledge Foundation are trying to do" (Interview April 22 2016). Similar to the applications developed by Experimenters, this helps to personalize investigative stories by allowing users to check their local situation. However, it expands it by integrating a civic tech application (FragDenStaat) to enable users to request more information and do their own investigations. Traditional

[27] https://correctiv.org/en/investigations/nursing-homes/.

investigation for stories and facilitating users exist side-by-side and extend each other.

In its usage of data, Correctiv's investigations also emphasize the granularity and completeness of data, similar to Experimenters. However, different to Experimenters, granularity and completeness is rationalized here as *enabling (journalistic) investigations* rather than creating useful services for readers. In part, creating complete and granular data is a side-effect of Correctiv's focus on long-term investigations: it concentrates on a small number of broad subjects and big investigations for longer periods of time. By using a subject rather than a hypothesis as a starting point, journalists are gathering large databases to get an overview and conduct a more exploratory form of data analysis. At the same time, granularity and completeness also matters for Correctiv's mission to strengthen investigative journalism and educating citizens:

> You could understand it as freeing data. We take data that was already available and put it into a form that allows people to investigate and compare it. We are creating a research tool for other journalists and citizens who are interested in this topic. (Interview March 30 2016)

Correctiv's services are not primarily intended to function as 'decision-making tools', but as research tools supposed to enable others to do their own investigations. This way, Translators are again standing between facilitating and gatekeeping, emphasizing granularity and completeness as a basis for story-driven gatekeeping.

Facilitators

> We're in favour of a vibrant, healthy, lively democracy. That means a rude and obnoxious place. Although we don't want to do that ourselves, it's entirely appropriate that we should facilitate other people to. (Steinberg quoted in Krotoski 2010)

As the civic tech scene is evolving and includes a very diverse set of actors (activists, corporations, governments), I will concentrate on mySociety, a non-profit organization from the UK founded in 2003 (Baack 2018). As mentioned above, mySociety is one of the oldest and most influential civic tech organizations today and exemplifies typical characteristics of civic tech.

Members of mySociety clearly position their work at the 'facilitating' end of the spectrum: "we don't want to get in the way. We don't want to be the gatekeeper" (Interview August 21 2015). They do not want to influence policy making or shape public debates, but provide the means that enable others to take action more effectively. The tools mySociety develops – for example problem reporting websites, parliamentary monitoring tools or freedom of information websites – are understood as services that empower citizens by letting them "see and be able to do what they are legally entitled to as easily as possible" (Interview August 21 2015). Most mySociety projects are intended to act like a 'layer' that translates the bureaucratic and legal procedures followed by public institutions into user-friendly interfaces and accessible language.

Members describe their role as being "a tool supplier for other organizations" (Interview August 17 2015). They reject calling their work 'journalism' and most of them do not have close ties to journalists. Overall, mySociety's relationship with journalism is multifaceted. Depending on the project and the

circumstances, journalists can be users of mySociety's tools, customers of its (technical) services, collaborators and partners, or all of these roles at once (see YourNextMP example above). Beside the fact that mySociety is not trying to tell stories, one of the biggest difference members see to journalism is the 'time scale' in which they work. Ultimately, mySociety is interested in creating a more participatory culture (Jenkins 2006) by making civic engagement easier and less time consuming for citizens. As mySociety's former CEO put it, this requires changing "what is normal rather than what is exceptional" (Interview August 17 2015). Part of mySociety's efforts is therefore to make users 'feel' empowered in a psychological sense: it's tools are intended to give them a "sense of agency" (Interview June 03 2015) by making engagement frictionless and demonstrating that their actions have an impact (Baack 2018). What mySociety is trying to achieve can be described as having 'long-term impact': a relatively slow change in perception of what is considered to be 'normal' about civic engagement. This requires its services to be popular and reliable. For this reason, mySociety thinks of itself to be more similar to companies like eBay, Amazon or TripAdvisor than to news media companies.

However, while mySociety sees its own work as distinct and different, the work of journalists is also seen as highly complementary to its own mission. This is best illustrated by 'Alaveteli Professional', a version of its freedom of information platform specifically designed for journalists. In the announcement, mySociety describes the role of journalists as complementing its mission of facilitating engagement:

> Citizen empowerment doesn't always happen by direct interaction with institutions. Feeling empowered and capable of affecting what happens in your community requires knowing what's going on in your community (…) it's hard to imagine a future in which ordinary people can be well-informed, without specialists doggedly

> asking questions of power, putting information from different sources together, and helping make sense of what's going on. (Crow 2016)

The (simplified) division of labor from mySociety's point of view correlates with the core categories identified here: mySociety as a civic tech organization facilitates citizens and professionals like journalists to take action themselves, while journalists inform the public about relevant events. mySociety primarily thinks of journalism as traditional watchdog journalism that acts as a gatekeeper to publicly relevant information. Normalizers identify with this ideal the most. However, when we take a closer look at the others groups within the figuration described here, we get a more nuanced picture.

With their emphasis on facilitating others through granular and complete data, both Experimenters and Translators are more similar to mySociety. The division of labor in these cases is more along the lines of applying 'generic' tools and mechanisms of acquiring information or getting in contact with authorities (mySociety) vs the 'practical application' of these tools and mechanisms in specific investigations. mySociety usually focuses on a small set of projects that should be generic and customizable to cover a broad range of use cases. Its use of data is therefore driven by a desire to scale technological solutions through standardization and reusability. By contrast, the extent to which tools and data formats can be standardized and made reusable in journalism (across all groups) is limited because "every investigation is different (…) has a different subject, requires different data, and has a different output" (Interview April 22 2016).

Correctiv's attempts to combine civic tech applications and investigative journalism and its relatively long-term focus on particular subjects creates expanded opportunities for sustained collaborations in joint projects with civic

tech organizations like mySociety. It seems likely that the collaborations between civic tech and journalism will primarily happen around specialized media organizations that follow the ProPublica or other non-profit models (Konieczna and Powers 2017) and do not focus on daily news reporting.

Conclusion

This article showed how data journalists and civic technologists complement each other in numerous ways because they overlap in their practical skills and aspirations. It showed how they form a figuration that exists along a shared continuum that oscillates between practices of facilitating and gatekeeping. Differences in how these axes are weighted result in different groups or 'articulations' within this continuum. Looking across organizational backgrounds and institutional settings, the individuals populating this community of practice can be described as *flexible data professionals* who aspire to work in a public interest: they share transferable skills in dealing with data and using information rights, and they want to use these skills to create public services that hold powerful people and institutions accountable and empower citizens.

The findings illustrate how practices of facilitating and of gatekeeping *complement* each other. What this article does *not* show is a weakening of gatekeeping in favor of open participation. First, while there are groups within the field of journalism with a stronger emphasis on facilitating, the majority of the interview partners were Normalizers who work in leading national newspapers. Together, they have a lot more authority to define and delineate 'journalism' than the other groups, who work in more experimental and sometimes also more precarious environments. Second and more importantly, those who emphasize facilitating do not oppose gatekeeping. On the contrary:

they think of their own work as facilitating not just ordinary citizens, but also professional gatekeepers, best illustrated by mySociety's 'Alaveteli Professional'. By taking the potential for related gatekeeping into account, actors with an emphasis on facilitating contribute to a more transparent and participatory form of gatekeeping. On the other hand, actors who emphasize gatekeeping also take related opportunities for facilitating into account to allow readers to explore the reported issue and learn how it affects them personally.

Rather than thinking of gatekeeping and facilitating in terms of competing 'logics' (Lewis 2012b) or some modern reincarnation of the Lippmann-Dewey debate (Schudson 2008), this article shows how the ongoing datafication of social life allows them to exist along a shared continuum and *mutually reinforce* each other. The overlap in practical skills and social imaginaries helped making journalism as a professional practice more permeable to outsiders and allowed actors outside the field of journalism to increasingly engage in practices traditionally attributed to journalism, as Rusbridger noted (see introduction). At the same time, data journalists move closer to civil society actors like civic technologists and complement their work.

These findings illustrate how datafication creates new entanglements between the field of journalism and civil society. To understand how datafication affects public knowledge production and the assemblage of publics, we should not only study how the reliance on data changes individual actors or fields such as journalism. We also need to ask what connections and entanglements datafication enables and what the implications of those entanglements are for the actors involved and for increasingly datafied publics.

7. Conclusions

Studying pioneer communities provides us with an opportunity to examine datafication not as an abstract process which "happens to us" (Hepp 2016, 927), but as a contested social transformation that is driven by the practices and social imaginaries of diverse groups – not just big corporations or governments, but also civil society actors and journalists. Amidst concerns that processes of datafication may threaten to undermine the agency of publics (cf. Zuboff 2015), this study shows how actors who aspire to work towards a public interest aim to use data to promote visions of citizenship, participation and government accountability.

In this thesis, I focus on two communities within civil society and journalism who are currently pioneering the use of data in order to assemble visions of democratic publics: activists in the open data and civic tech movements and data journalists. I address two research questions: (1) *What is the role of data in the social imaginaries and practices of data activists and data journalists*; and (2) *how do the practices and imaginaries of these actors diverge and converge, and how does this shape the entanglements between them?* By providing answers to these questions, this thesis aims to complement a growing body of research which insists that a complex transformation of social life is taking place, in which the agency of publics does not simply erode due to new power structures around datafication, but where new forms of agency are emerging (Couldry, Fotopoulou, and Dickens 2016).

Chapter 2 reviews a selection of research literature on three aspects relevant to the understanding of data journalists and data activists as pioneer communities: a) the historical roots of data journalism and data activism, b)

the relationship and direct interactions between these two groups, and c) the implications of these phenomena, i.e. their potential causes and effects. I discuss how data journalism and data activism can be understood as related to broader developments within democratic practice and discourse: the evolution from representative to monitorial forms of democracy, and a related transformation from informed to monitorial citizens (Schudson 1998, 2015; Keane 2009). I argue that the premises of monitorial democracy – growing demands to scrutinize governments and for more continuous forms of representation – are related to datafication. Due to the fact that practices of quantification are essential for evaluating governments and ensuring representativeness, such demands are likely to make practices of quantification more important and widespread. Suggesting that commercial or government interests are the only relevant drivers of datafication overlooks both the historical and contemporary connections of datafication to non-profit and non-governmental interests. I also point to two of the main shortcomings of the current research literature for understanding data journalists and data activists as pioneer communities. First, a critical in-depth examining of the practices and imaginaries of data activists is missing thus far. Often, researchers tend to implicitly or explicitly assume that the practices and values of activists in the open data and civic tech movements can be equated with open source culture, which ignores that open source culture is being modulated and transformed when it is applied to new domains. Second, research on the relationship between data activists and data journalists narrowly focuses on the direct interactions between these groups. Rather than examining how data journalists and data activists are able to complement each other, the current literature is largely focused on how open source culture is

integrated in newsrooms, or how technologists and journalists are able to collaborate and develop cross-understanding.

Chapter 3 describes the methodological approach of this study. By combining the focus on practices with a constructivist grounded theory approach relying on interviews, content analysis, and participatory mapping, the aim of this thesis is to be sensitive to nuances regarding the ways in which data activists and data journalists understand their work, the role of data within that work, and how data activists and data journalists relate to each other. I conducted three case studies: two studies focus on 'best practice' non-profit civic tech organizations, and the third one explores the entanglements between data activists and data journalists.

In this conclusion, I first summarize how the findings described in Chapters 4-6 address my research questions. Next, I discuss the methodological and theoretical implications of this study by returning to the main concepts and theoretical frames outlined in Chapter 2. I end with arguing that future research on the implications of datafication for democratic publics requires both empirical research and normative theory.

Summary of the findings

1. What is the role of data in the social imaginaries and practices of data activists and data journalists?

Collectively, the findings discussed in Chapters 4-6 show that data journalists and data activists do not invent new social imaginaries around data. Instead, the ways that they use and understand data follow notions of participatory culture and journalistic gatekeeping, which both pre-date our modern, digital

media environment. Nonetheless, data activists and data journalists are pioneering the use of data to implement these imaginaries in new ways.

In Chapters 4-5, I show how activists in the open data and civic tech movements imagine and attempt to realize a 'data-driven' form of participatory culture, in which data is used to facilitate citizens to engage with authorities and become more active and involved. With services such as freedom of information websites or tools for parliamentary monitoring, data activists aim to make engagement with authorities less time consuming and more 'user-friendly' from the perspective of citizens. By facilitating greater engagement with governments, data activists want to ensure that decision-making "reflects the pooled inputs of a distributed population" (Lewis 2012b, 847). In concord with this mission, data activists understand themselves as facilitators of civic engagement. This means that they rarely advocate for specific policies themselves, but aim to enable others to advocate their causes more effectively. In other words, data activists want to influence the conditions by which others participate in the public arena without directing the public discourse or influencing policy making. At the same time, they see their role as complementing actors who *do* want specific outcomes (e.g. specific policy changes) by facilitating advocacy groups within civil society or professional journalists. Data forms an important basis for the broader imaginaries of activists, and it is key to how those activists attempt to realize their imaginaries.

In Chapter 4, I show how open data activists combine participatory culture with data by applying practices and values from open source culture to the creation and use of data. First, open data activists treat raw data as a form of 'source code', whose interpretation and re-use generates knowledge. Sharing raw data openly (analogous to how open source shares software code) would

allow citizens to make their own interpretations of the data collected by governments in order to make and justify decisions. Second, on the basis of this 'data transparency', open data activists apply the open source model of voluntary participation (cf. Weber 2004) to civic engagement and political participation. Voluntary participation is about designing production- or decision-making processes in ways that allow potentially everybody to participate in a meaningful way. Postulating that information is a prerequisite for participation, open data activists want governments to include interested citizens in decision-making processes on the basis of data transparency. This emphasis on citizen-involvement leads them to develop a vision of a more 'open' and flexible form of representative democracy, which has a greater level of transparency and more continuous forms participation beyond periodic voting. While such a vision is not new in itself (cf. Barber 2004), the way that data is imagined to enable this vision is novel. Chapter 4 also shows how data is used to realize this imaginary, as activists aim to create and act as 'empowering intermediaries' whose function is to make raw data accessible to the wider public.

In Chapter 5, I expand the findings described in Chapter 4 by critically examining how members of the British NGO mySociety use data to realize their imaginaries and how they position themselves within the public arena. I illustrate that members of mySociety use and understand data in ways that are supposed to expand the agency of publics towards governments. Most importantly, the re-use and creation of data allows mySociety to mediate between governments and their publics. The tools of mySociety are designed to function as 'layers' between citizens and the legal and bureaucratic procedures followed by government institutions. On the one hand, the databases created by mySociety make the performance of governments legible

to the public in new ways. For example, mySociety's problem-reporting website, FixMyStreet, collects and publishes all of the issues that users report to local authorities (e.g. potholes or broken streetlights), and tracks whether or not those issues are being resolved. This surveillance pressures authorities to cooperate with mySociety's tools, and adapt mySociety's more 'user-oriented' ways of providing public services or allow citizens to make use of existing rights. On the other hand, members of mySociety attempt to use databases to connect the individual with the collective. They try to demonstrate to the users of their applications that their actions have an impact and create a public value, e.g. by showing that reporting local issues via FixMyStreet helps to resolve issues in one's neighborhood. In this way, data is imagined to change the perception of users about how much power they have, and what they consider to be 'normal' about engaging with authorities.

Using data to influence both governments and citizens, mySociety essentially re-interprets how existing legal frameworks and public services should 'ideally' work, and tries to impact the ways in which they are being implemented. The applications and services of mySociety remain compatible to the workings of government, but are intended to highlight dissonance between the way institutions implement laws and public services and the way to ideally implement these laws and services from a citizen's point of view. Additionally, mySociety uses structured data to make 'deep linking' into documents possible. Deep linking is about making documents searchable, and attributing speeches to particular individuals. As the processes of governments are largely captured in documents, e.g. parliamentary speeches, deep linking aims to increase the publics' awareness of, and engagement with, the processes of governments. Finally, mySociety promotes data standards and standardized tools in order to support a distributed form of agency, which should enable groups in other

countries who similarly seek to promote a more participatory culture to develop tools that serve their particular needs.

In Chapter 6, I examine the practices and imaginaries of data journalists. I show that the ways in which data journalists understand and imagine data varies depending on their educational background and the organizational settings that they are working in. We can roughly distinguish two main practices. First is the use of data to support traditional journalistic gatekeeping. Gatekeeping entails selecting and communicating publicly relevant information. Actors who put the emphasis on gatekeeping aim to direct and steer public debate by highlighting issues they consider to be relevant. Therefore, their use of data is story-driven. What matters is the quality of the information, i.e. its ability to prove or falsify a hypothesis, or to reveal the scope of a public issue (cf. Parasie and Dagiral 2013). Second is the use of data to facilitate others to take action themselves – a practice that has, arguably, been pioneered by data activists, but which can also be found among data journalists. Data activists and data journalists who emphasize facilitating are concerned with the quantity and structure of data. For them, data should be granular and complete so it can be broken down to serve individual users. For example, the data underlying an application should allow users to enter their address to acquire individualized information, e.g. the quality of their local tap water or their representatives in parliament.

Beside data activists, I identify three groups of actors who work within the field of professional journalism, but differ in the level of emphasis that they place on gatekeeping and facilitating: Normalizers, Experimenters and Translators. Normalizers are data journalists with a formal education in journalism, who are working in established national news media companies. Of the three identified groups, Normalizers have the strongest emphasis on

gatekeeping, and use data in ways that serve long-standing notions of investigative, professional journalism. Data is used and understood by Normalizers as a means to conduct journalism the way it has been conducted for decades, but more efficiently and on a larger scale. By contrast, Experimenters and Translators combine gatekeeping and facilitating – which brings me to my second research question.

2. How do the practices and imaginaries of data activists and data journalists diverge and converge, and how does this shape the entanglements between them?

Understanding the ways in which practices of facilitating and gatekeeping complement each other is key to understanding the entanglements between data activists and data journalists. The findings that I present in Chapter 6 show that the boundaries between data activism and data journalism are blurry: some journalists adopt practices and values similar to those of data activists and combine them with journalistic gatekeeping. I identify three themes which show where and how data journalists and data activists come together: overlapping and transferable skills necessary to work with data, a commitment to learning and open source culture, and complementary ambitions. In short, data activists and data journalists rely on a similar skill set to work with data, enjoy tinkering with technology and sharing (code, data, or experiences), and aspire to provide a public service that holds powerful people or institutions accountable. They easily conceive each other as complementary, which occasionally allows seamless transitions between projects by data activists and data journalists. The practices of data journalists and data activists 'interlock' in this sense and the actors form a community of practice or a figuration, i.e. a network of human actors based on interlocking practices and shared meanings (Couldry and Hepp 2017, Chapter 4).

The work of Translators best illustrates how facilitating and gatekeeping can interlock and mutually reinforce each other. Translators are active in both journalism and civic tech, and work in news media organizations at the intersection of both fields. In Chapter 6, I discuss the example of Correctiv, a nonprofit newsroom which strives to simultaneously strengthen investigative journalism and teach citizens to become journalists themselves. Like Normalizers, Correctiv aims at steering public debate through gatekeeping, but it also strives to enable the users of their tools to work like journalists and conduct their own investigations. Civic tech applications help to bridge these goals because they function as research tools that can be useful for both professional journalists and ordinary citizens. Often, Correctiv's projects begin with a particular subject, e.g. the nursing home sector in Germany. The data that Correctiv collects is granular and complete in order to enable others to explore it themselves. At the same time, what Correctiv aims to facilitate is not the engagement of ordinary citizens with their governments, but journalistic gatekeeping. Gatekeeping and facilitating should mutually reinforce each other: highlighting important issues via gatekeeping should raise awareness and interest in the subject. Interested readers or other journalists are then invited to explore the highlighted issue themselves with the help of the tools provided by Correctiv, in hopes that this exploration will help to uncover new issues, which can then be further highlighted via gatekeeping.

In similar ways, gatekeeping and facilitating complement each other across organizational boundaries. While Normalizers emphasize differences between their own work and those of data activists, they are interested in exploring opportunities to complement their reporting with tools that enable their audiences to explore the issues they report on. At the same time, data activists are interested in making their tools attractive to journalists, and investigations

conducted by journalists have sparked ideas for new civic tech applications (e.g. 'KleineAnfragen.de' mentioned in Chapter 6). The last group, Experimenters, introduces another variation of how gatekeeping and facilitating are being understood and weighted against each other. Experimenters often have a background in the technology sector and tend to question the boundaries and meanings of professional journalism. Even though Experimenters work within media organizations, they emphasize facilitating over journalistic gatekeeping. Still, they consider their work to be an expansion of journalism. According to Experimenters, journalists should tell stories, but also provide readers with a new access to their surroundings. Similar to data activists, they therefore aim to enable their users to explore issues themselves. The main difference between Experimenters and Translators is that Experimenters usually do not wish to facilitate other journalists to do their own investigations.

As discussed in Chapter 2, participatory culture and professional journalistic gatekeeping are often perceived as different and competing 'logics' that negate each other: either one ensures open participation or professional control (cf. Lewis 2012b). My findings challenge this assumption. The growing reliance on data allows practices of facilitating and gatekeeping to interlock and exist along a continuum. Data is used in ways that both strengthen journalistic gatekeeping and facilitate others to take action themselves. This not only has important implications for (the entanglements of) data activists and data journalists, but also for how we should theorize and study them.

Implications

In comparing my findings with the theoretical frameworks outlined in Chapter 2, I want to make two points. First are the methodological implications. This

thesis illustrates the need to be more sensitive to the ways in which processes of datafication are related to new, and emerging figurations, i.e. to new networks of human actors that cross institutional or organizational boundaries. This is particularly important in the light of recent and ongoing developments in journalism. Building on those methodological suggestions, I outline some theoretical implications. The findings of this study demonstrate that we should be careful in how we apply existing theoretical concepts to explain ongoing transformations around datafication. While we can understand data activists and data journalists in terms of monitorial democracy and participatory culture, solely relying on these concepts can be problematic and we need more theory-building. Finally, I argue that the study of datafication should strive to bring empirical research and normative theory together by using the concept of agency.

Methodological implications

In Chapter 2, I show that much of the existing research about data journalists or data activists considers each group in isolation. Journalism studies, especially, still tends to have a very 'newsroom-centric' perspective. This is not only in the sense that it is primarily interested in how data journalism is practiced within newsrooms, but also in the ways it tends to reproduce the professional boundaries of journalism and conduct "back-up boundary work" (Wahl-Jorgensen 2014, 2588; cf. Domingo, Masip, and Costera Meijer 2015). For example, studies rely on concepts such as 'trading zones' (Lewis and Usher 2014) or 'boundary objects' (Lewis and Usher 2016) to examine how journalists and technologists cooperate and manage to achieve mutual understandings in workshops or in joint projects. Research on the relationships between journalists and technologists with roots in open source

culture usually asks how, where and why the professional logic of control becomes "rearticulated (or not) in relation to the participatory logic" (Lewis 2012b, 852). This work is important, but I argue that we must also explore how journalists and other actors are able to complement each other *without* adapting each other's practices and values, which is particularly relevant given that open source advocates are able to build recursive publics that can exist independently, outside of established organizational or institutional structures (see Chapter 2 and 4). How does the long-term co-existence, the continued exchange and mutual awareness affect how data journalists and data activists understand their work and how they utilize data?

By moving away from a firm distinction between data journalists and data activists, this study provides a more nuanced picture of how these actors complement each other. I show how the shared reliance on data and complementary ambitions allow data journalists and data activists to form a figuration across diverse organizational and institutional backgrounds. These findings show that the growing reliance on data creates new links between actors in key areas of public space: journalists and civil society actors. If we are to understand datafication's influence on democratic practices, we need to be sensitive to the ways in which processes of datafication are connected to such emerging figurations. As Couldry (2012, 57) puts it, "new figurations are emerging around us, but it may be some time before their shape is clear". This has important implications for how we study and understand data journalism and data activism.

On the one hand, the results of this study show that a focus on how professional journalists integrate and 'normalize' (Singer 2005) the practices and values of data activists is too narrow. While normalization makes perfect sense for one group identified ('Normalizers'), there are also types of actors

within the field of journalism who cannot be grasped within the framework of 'traditional journalism': Experimenters and Translators. In their practices, self-understanding, and use of data, these actors are much closer to data activism than Normalizers and they do not necessarily identify with traditional notions of journalism. They do not rely on strong traditional routines and identities and consequentially may not have a clear reference point on how practices 'external' to their work could be normalized. These findings reflect the ongoing transformation of journalism, whose "rich and relatively stable history of professionalization" (Deuze and Witschge 2017, 1) seems to have reached a turning point. The field of journalism is transitioning from a "more or less coherent industry to a highly varied and diverse range of practices" (Deuze and Witschge 2017, 2). Both Experimenters and Translators occupy 'non-traditional' spaces of work within the field of journalism. Similar to the findings of Deuze and Witschge (2017, 13), this thesis suggests that journalism studies would benefit from broadening its focus, and understanding journalism as a "dynamic set of practices and expectations – a profession in a permanent process of becoming".

On the other hand, this thesis shows that we also need to take the role of (data) journalists into account in order to understand data activists. First, the idea to support the work of professional journalists with their tools is an important part of how data activists self-identify as facilitators of engagement. Second and related to the first point, some data activists place importance on journalists using the tools that they provide, as this is perceived as a way to put their ideas into practice, and potentially have a larger impact. This was most visible in the motives of Translators described in Chapter 6. Third, transforming of journalism to be more data-driven and participatory is a key element in the social imaginaries of at least some data activists, who have made

teaching their methods and techniques to journalists an important part of their mission (see Chapter 4).

The methodologies employed in this study are particularly helpful to identify and examine emerging figurations, as they enable the researcher to be sensitive to how actors understand their work, to how they position themselves professionally, and to their subjective sense of belonging. As the practices and imaginaries studied here are developed by pioneer communities, the next step would be to investigate how these practices and imaginaries are being adapted, and how this is creating or changing figurations within and between civil society and journalism. Based on the research that I have presented here, I submit that these three aspects should be taken into account to critically examine the implications, and to develop an analytical sensibility to how a more widespread adaption is reinforcing, or shifting existing power structures:

1. *The aims and imaginaries of the actors involved: what do they want to facilitate and how do they want to expand gatekeeping?* Do the actors involved wish to change the way certain rights can be used by citizens, or do they prefer to facilitate professional gatekeepers to more effectively monitor authorities? Practices of facilitating will expand gatekeeping and distribute agency in distinct ways, depending on what kinds of actions actors aim to facilitate and who they consider to be the primary users. At the same time, how do actors interested in gatekeeping take related opportunities to facilitate users into account? What are the implications of the resulting dynamics between facilitating and gatekeeping?

2. *How facilitators act as intermediaries: how do they attempt to change existing processes?* In Chapter 5, I show how data activists mediate between the actual processes of governments and their own imaginaries. If we are to

understand the implications of facilitating, then we ought to ask how actors who aim to facilitate others imagine a process should ideally work, and how they attempt to realize their imaginaries using data. What are the differences between the existing processes and their visions, and how do these differences reflect on wider imaginaries? For example, a problem-reporting website such as FixMyStreet might appear mundane at first glance, but it is part of mySociety's broader vision of a more participatory culture. FixMyStreet does not only aim to make problem-reporting easier and less time-consuming, but also imagines problem-reporting as a public process, which enables a more collaborative relationship between authorities and citizens. The ways in which broader imaginaries are translated into the design of tools and the use of data are not always obvious; they often require careful investigation into the sense-making processes of the actors involved.

3. *Existing power structures: are they taken into account and if so, how?* Research has demonstrated numerous times that developing tools to facilitate engagement with authorities without considering questions of inequality and power will merely reinforce existing structures and 'empower the empowered' (Gurstein 2011). Are facilitators sensitive to dynamics of inequality and power, and how does this influence their work? What are the consequences if facilitators do not take such dynamics into account?

As I discuss in the next section, addressing these questions will also invite us to critically examine the theoretical concepts that we currently rely on to grasp new figurations.

Theoretical implications

In Chapter 2, I discuss how data activists, data journalists, and the entanglements between them have been theorized in terms of monitorial democracy and open source culture. The findings in Chapters 4-6 confirm that these concepts are useful for understanding the practices and imaginaries of these actors. However, as data journalists and data activists implement and combine notions of monitorial democracy and participatory culture in new ways, we cannot fully grasp the figuration they form if we solely study them through the analytical lenses these concepts provide. Monitorial democracy and open source or participatory culture presuppose particular images of citizenship, but the ways in which practices of facilitating and gatekeeping complement each other allows for flexibility in the accommodation of different types of engagement and, accordingly, different types of citizenship. Rather than equating the practices and social imaginaries of data activists and data journalists with particular models of democratic engagement, we ought to be sensitive to how the figurations formed by actors in the field of civil society and journalism combine different notions of citizenship.

Allow me to illustrate this point for practices of facilitating. Through such practices, data journalists and data activists combine and, by way of combining them, re-articulate notions of participatory culture and monitorial democracy. One might understand facilitating as a way to conciliate monitorial democracy and participatory culture: by using data in ways that enable others to take action themselves, data journalists and data activists attempt to simultaneously enable their users to monitor the performance of governments (e.g. via parliamentary monitoring websites) and to be more active and engaged. This idea would also support the argument made in Chapter 2: monitorial

democracy and participatory culture are related phenomena because both intent to ensure that a multiplicity of voices are present in decision-making processes, within and beyond the institutionalized practices of representation. However, the findings made in Chapters 4-6 show that it is problematic to conceptualize facilitating in those terms because monitorial democracy and participatory culture contradict each other in crucial ways.

First, participatory culture is about creating an active, engaged, and collaborative community (Lewis 2012b). In contrast, Schudson's (1998, 311) monitorial citizens are "defensive rather than pro-active" and only take actions if an issue arises which demands their intervention. While the facilitation conducted by data activists and data journalists often involves monitoring, practices of facilitating ultimately inherit what Lewis (2012b, 848) describes as the normative imperative of participatory culture: the ambition to create "a more engaged, representative, and collectively intelligent society". Still, by making monitoring authorities easier, data journalists and data activists typically accommodate both active and passive citizens. Civic tech applications, such as parliamentary monitoring websites, will usually offer services to both occasional users and (professional) power-users, and data journalists will seek opportunities to facilitate their audiences based on the information they provide. In short, facilitating can be attributed to both monitorial democracy and participatory culture, but we cannot fully understand these practices if we solely rely on one of those concepts.

Second, the ways in which facilitating complements gatekeeping do not align easily with monitorial democracy or participatory culture. Monitorial democracy suggests that we have entered an era of 'post-representational' forms of citizenship, where the classic ideal of the informed citizen has become less prominent (Keane 2009). Yet, facilitating complements

journalistic gatekeeping, which continues to rely on informed models of citizenship (cf. Schudson 1998). Additionally, participatory culture is about open and collaborative production, which is conventionally understood as problematic for journalistic notions of professional control and gatekeeping (Lewis 2012b). More fundamentally, participatory culture is connected to a political philosophy that Tkacz (2012) describes as 'open politics'. Of central importance for open politics is "whether or not a social programme, that is, a set of knowledges and related practices, is able to change" (Tkacz 2012, 389). It is not geared towards specific change, but "towards change in general" (Tkacz 2013). However, we cannot understand facilitating in those terms because practices of facilitating are intended to complement others who want specific outcomes, be it ordinary citizens engaging with their governments or professional journalists who are interested in finding stories.

The difficulties to fit practices of facilitating into existing theoretical concepts illustrates the numerous ways in which new figurations may connect distinct democratic visions. To grasp emerging figurations around datafication, we should consider how we are using existing concepts. Established concepts do not have to become obsolete, but if we solely rely on them, they may hinder our understanding of new developments within civil society and journalism. By identifying and theorizing the figurations formed between data activists and data journalists, I have aimed to contribute to the broader task of revisiting and revising the theoretical toolkit that is necessary for understanding the assemblage of datafied publics.

Looking ahead: Agency in datafied publics

Going forward, a critical examination of datafication will require both empirical research and normative theory. I argue that the concept of agency, understood as reflexive practice (see Chapter 4), is key for bringing those perspectives together. Questions of agency are inherently normative; to ask who has how much agency is to ask about power relationships, and such questions are increasingly related to processes of datafication. Moreover, normative questions are relevant to the study pioneer communities given their anticipated influence. As I argue in Chapter 1, if we want to assess the implications of a more widespread adoption of the practices and imaginaries of pioneer communities, we must take the power relationships embedded in those practices and imaginaries into account. While this study is not primarily about agency itself, I have raised questions about agency throughout. I show how data expands the agency of data activists and data journalists, and ask about the implications for the publics they aim to assemble.

In this way, I provide insight into the normative assumptions that guide these actors in the ways in which they appropriate data. Further research into similar practices should be complemented by research that explicitly takes a normative perspective, and asks how datafication ought to be regulated in order to "enlist processes of datafication into the service of social progress," as Gray (2016) eloquently puts it. Given that the technologies, institutions, and everyday practices that drive datafication are not going to disappear, this research explores the possibilities to mitigate the negative consequences of datafication and enable us to utilize processes of datafication to support agency and democratic governance in various ways. As Kennedy and Moss (2015, 5) ask:

are "alternative data practices and arrangements (…) possible and, if so, what [should] they (…) look like"?

Empirical research into the practices and imaginaries of pioneer communities can inform both such normative debates and practitioners themselves. Researchers interested in normative questions and pioneer communities who aspire to work in a public interest hold similar goals: preserving or expanding long-standing normative ideals of democracy, equality, and accountability. Empirical research can act as bridge between normative theory and actors who use data to serve a public interest, and facilitate a dialog. As argued in Chapter 5, the democratic possibilities of data cannot be determined in an abstract way because what data affords to whom is historically and culturally situated, and therefore need to be qualitatively contextualized. To understand what a growing reliance on data means for democracy, we must develop a qualitative sensitivity for how quantification affects the values we aim to preserve or expand. Throughout this thesis, I have aimed to illustrate how such a qualitative approach can be conducted, and hope to have contributed to an ongoing discussion about the implications of datafication for the collectives that we form.

References

Alasuutari, Pertti. 1996. "Theorizing in Qualitative Research: A Cultural Studies Perspective." *Qualitative Inquiry* 2 (4):371–84. https://doi.org/10.1177/107780049600200401.

Ananny, Mike, and Kate Crawford. 2015. "A Liminal Press." *Digital Journalism* 3 (2):192–208. https://doi.org/10.1080/21670811.2014.922322.

Anderson, C. W. 2011. "Deliberative, Agonistic, and Algorithmic Audiences: Journalism's Vision of Its Public in an Age of Audience Transparency." *International Journal of Communication* 5:529–47. http://ijoc.org/index.php/ijoc/article/view/884/537.

Anderson, C. W. 2013a. *Rebuilding the News: Metropolitan Journalism in the Digital Age*. Philadelphia: Temple University Press.

Anderson, C. W. 2013b. "Towards a Sociology of Computational and Algorithmic Journalism." *New Media & Society* 15 (7):1005–21. https://doi.org/10.1177/1461444812465137.

Anderson, C. W. 2015. "Between the Unique and the Pattern." *Digital Journalism* 3 (3):349–63. https://doi.org/10.1080/21670811.2014.976407.

Appelgren, Ester, and Gunnar Nygren. 2014. "Data Journalism in Sweden." *Digital Journalism* 2 (3):394–405. https://doi.org/10.1080/21670811.2014.884344.

Arcopix. 2014. "Re-Usable Components for Civic Tech - Dave Whiteland." https://www.youtube.com/watch?v=2v4tVpPwIs4\&gt.

Baack, Stefan. 2013. "Die Open-Data-Bewegung. Das Verhältnis von Praktiken, Zielen und Selbstbild der Open Knowledge Foundation Deutschland." University of Bremen. http://nbn-resolving.de/urn:nbn:de:0168-ssoar-363745.

Baack, Stefan. 2015a. "Datafication and Empowerment: How the Open Data Movement Re-Articulates Notions of Democracy, Participation, and Journalism." *Big Data & Society* 2 (2). https://doi.org/10.1177/2053951715594634.

Baack, Stefan. 2015b. "Scraping the Global Civic Tech Community on GitHub, Part 2." *Civic Hacking and Journalism*. http://sbaack.com/2015/11/19/scraping-the-global-civic-tech-community-on-github-part-2.html.

Baack, Stefan. 2016. "What Big Data Leaks Tell Us About the Future of Journalism - and Its Past." *Internet Policy Review*. https://policyreview.info/articles/news/what-big-data-leaks-tell-us-about-future-journalism-and-its-past/413.

Baack, Stefan. 2017. "Practically Engaged: The Entanglements Between Data Journalism and Civic Tech." *Digital Journalism*, October. https://doi.org/10.1080/21670811.2017.1375382.

Baack, Stefan. 2018. "Civic Tech at mySociety: How the Imagined Affordances of Data Shape Data Activism." *Krisis* 2018 (1).

Bachmann, Götz, and Andreas Wittel. 2011. "Medienethnographie." In *Qualitative Methoden der Medienforschung*, edited by Ruth Ayaß and Jörg Bergmann, 183–219. Mannheim: Verlag für Gesprächsforschung.

Baraniuk, Chris. 2013. "Power Politechs." *New Scientist* 218 (2923):36.

Barber, Benjamin R. 2004. *Strong Democracy: Participatory Politics for a New Age*. Berkeley, CA etc.: University of California Press.

Barocas, Solon, and Andrew Selbst. 2016. "Big Data's Disparate Impact." *California Law Review* 104 (3):671. https://doi.org/10.15779/Z38BG31.

Bates, Jo. 2012. "'This Is What Modern Deregulation Looks Like': Co-Optation and Contestation in the Shaping of the UK's Open Government Data Initiative." *The Journal of Community Informatics* 8 (2). http://ci-journal.net/index.php/ciej/article/view/845.

BBC. 2003. "Hunt for 'Napster of Good Causes'," October. http://news.bbc.co.uk/2/hi/technology/3228339.stm.

Bell, Emily. 2014. "Silicon Valley and Journalism: Make up or Break up?" *Reuters Institute for the Study of Journalism*. http://reutersinstitute.politics.ox.ac.uk/news/silicon-valley-and-journalism-make-or-break.

Benson, Rodney, and Erik Neveu. 2005. "Introduction: Field Theory as a Work in Progress." In *Bourdieu and the Journalistic Field*, edited by Rodney Benson and Erik Neveu, 1–25. Cambridge ; Malden, MA: Polity.

Berners-Lee, Tim. 2010. "Linked Data." http://www.w3.org/DesignIssues/LinkedData.html.

Beyer, Jessica L. 2014a. "The Emergence of a Freedom of Information Movement: Anonymous, WikiLeaks, the Pirate Party, and Iceland." *Journal of*

Computer-Mediated Communication 19 (2):141–54. https://doi.org/10.1111/jcc4.12050.

Beyer, Jessica L. 2014b. *Expect Us: Online Communities and Political Mobilization*. Oxford; New York: Oxford University Press.

Bounegru, Liliana. 2015. "GitHub as Transparency Device in Data Journalism, Open Data and Data Activism." http://lilianabounegru.org/2015/07/08/github-as-transparency-device-in-data-journalism-open-data-and-data-activism/.

Bowden, Tony. 2010. "Dream Out Loud." *Agitprop Àgogo. A Civic Tech Abecedary*. http://agitagogo.com/abecedary/d/.

Bowden, Tony. 2014a. "A Poplus Creation Story." *[Dis]compose*. http://discomposer.com/creation/.

Bowden, Tony. 2014b. "Harnessing Self-Interest." http://discomposer.com/harnessing-self-interest/.

Bowers, Thomas A. 1976. "'Precision Journalism' in North Carolina in the 1800s." *Journalism Quarterly* 53 (4):738–40. https://doi.org/10.1177/107769907605300422.

boyd, danah, and Kate Crawford. 2012. "Critical Questions for Big Data." *Information, Communication & Society* 15 (5):662–79. https://doi.org/10.1080/1369118X.2012.678878.

Breindl, Yana. 2016. "Activists as News Producers." In *The SAGE Handbook of Digital Journalism*, edited by Tamara Witschge, C. W Anderson, David

Domingo, and Alfred Hermida, 250–65.
http://dx.doi.org/10.4135/9781473957909.n17.

Brennen, Bonnie S. 2013. *Qualitative Research Methods for Media Studies*. New York, NY [u.a.]: Routledge.

Cantijoch, Marta, Silvia Galandini, and Rachel Gibson. 2016. "'It's Not About Me, It's About My Community': A Mixed-Method Study of Civic Websites and Community Efficacy." *New Media & Society* 18 (9):1896–1915. https://doi.org/10.1177/1461444815616225.

Carlson, Matt. 2015. "The Robotic Reporter." *Digital Journalism* 3 (3):416–31. https://doi.org/10.1080/21670811.2014.976412.

Charmaz, Kathy. 2006. *Constructing Grounded Theory*. 1. ed. London; Thousand Oaks, Calif: Sage Publications.

Clement, Andrew, and Christie Hurrell. 2008.
"Information/Communications Rights as a New Environmentalism? Core Environmental Concepts for Linking Rights-Oriented Computerization Movements." In *Computerization Movements and Technology Diffusion: From Mainframes to Ubiquitous Computing*, edited by Margaret S. Elliott and Kenneth L. Kraemer, 337–58. Medford, New Jersey: Information Today, Inc.
http://hdl.handle.net/1807/35202.

Coddington, Mark. 2015. "Clarifying Journalism's Quantitative Turn." *Digital Journalism* 3 (3):331–48. https://doi.org/10.1080/21670811.2014.976400.

Coleman, Gabriella. 2013. *Coding Freedom: The Ethics and Aesthetics of Hacking*. Princeton: Princeton University Press.
http://gabriellacoleman.org/Coleman-Coding-Freedom.pdf.

Coleman, Gabriella. 2014. *Hacker, Hoaxer, Whistleblower, Spy: The Many Faces of Anonymous*. 1 edition. London; Brooklyn, NY: Verso.

Coleman, Gabriella. 2016. "Hacker." In *Digital Keywords: A Vocabulary of Information Society and Culture*, edited by Benjamin Peters, 158–72. Princeton Studies in Culture and Technology. Princeton: Princeton University Press.

Coleman, Gabriella, and Alex Golub. 2008. "Hacker Practice: Moral Genres and the Cultural Articulation of Liberalism." *Anthropological Theory* 8 (3):255–77. https://doi.org/10.1177/1463499608093814.

CORRECTIV. 2015. "CORRECTIV-Schulung: Journalistisch Arbeiten - Was Ist Ein Journalist." https://www.youtube.com/watch?v=DnwtJyFT7JQ.

CORRECTIV. n.d. "Frequently Asked Questions." *CORRECTIV Homepage*. Accessed December 3, 2016. https://correctiv.org/en/correctiv/faq/.

Couldry, Nick. 2004. "Theorising Media as Practice." *Social Semiotics* 14 (2):115–32. https://doi.org/10.1080/1035033042000238295.

Couldry, Nick. 2010. *Why Voice Matters: Culture and Politics After Neoliberalism*. Los Angeles; London: Sage.

Couldry, Nick. 2012. *Media, Society, World: Social Theory and Digital Media Practice*. Cambridge; Malden, MA: Polity.

Couldry, Nick. 2014. "Inaugural: A Necessary Disenchantment: Myth, Agency and Injustice in a Digital World." *The Sociological Review* 62 (4):880–97. https://doi.org/10.1111/1467-954X.12158.

Couldry, Nick, Aristea Fotopoulou, and Luke Dickens. 2016. "Real Social Analytics: A Contribution Towards a Phenomenology of a Digital World." *The British Journal of Sociology* 67 (1):118–37. https://doi.org/10.1111/1468-4446.12183.

Couldry, Nick, and Andreas Hepp. 2017. *The Mediated Construction of Reality*. Cambridge, UK ; Malden, MA: Polity Press.

Couldry, Nick, and Alison Powell. 2014. "Big Data from the Bottom up." *Big Data & Society* 1 (2). https://doi.org/10.1177/2053951714539277.

Cox, Melisma. 2000. "The Development of Computer-Assisted Reporting." In. Phoenix. http://com.miami.edu/car/cox00.pdf.

Crabtree, James. 2003. "Civic Hacking: A New Agenda for E-Democracy." *openDemocracy*. https://www.opendemocracy.net/debates/article-8-85-1025.jsp.

Cridge, Mark. 2015. "Why We Do What We Do." *mySociety Blog*. https://www.mysociety.org/2015/11/17/why-we-do-what-we-do/.

Crow, Louise. 2016. "Alaveteli Professional Learning More About Journalistic Use of Freedom of Information." *mySociety*. https://www.mysociety.org/2016/09/06/alaveteli-professional-learning-more-about-journalistic-use-of-freedom-of-information/.

DataDrivenJournalism.net. 2016. "GitHub Experiment: Mapping the Civic Tech Community." *Data Driven Journalism*. http://datadrivenjournalism.net/featured_projects/github_experiment_mapping_the_civic_tech_community.

Davies, Tim. 2010. "Open Data, Democracy and Public Sector Reform: A Look at Open Government Data Use from Data.Gov.Uk." MSc Dissertation, University of Oxford. http://www.opendataimpacts.net/report/.

DeBarros, Anthony. 2010. "Data Journalism and the Big Picture." http://www.anthonydebarros.com/2010/11/26/data-journalism-the-big-picture/.

Demil, Benoît, and Xavier Lecocq. 2006. "Neither Market nor Hierarchy nor Network: The Emergence of Bazaar Governance." *Organization Studies* 27 (10):1447–66. https://doi.org/10.1177/0170840606067250.

Deuze, Mark. 2008. "The Changing Context of News Work: Liquid Journalism for a Monitorial Citizenry." *International Journal of Communication* 2 (July):18. http://ijoc.org/index.php/ijoc/article/view/290.

Deuze, Mark, and Tamara Witschge. 2017. "Beyond Journalism: Theorizing the Transformation of Journalism." *Journalism*, February. https://doi.org/10.1177/1464884916688550.

Dey, Ian. 1999. *Grounding Grounded Theory: Guidelines for Qualitative Inquiry*. San Diego: Academic Press.

Diakopoulos, Nicholas. 2011. "A Functional Roadmap for Innovation in Computational Journalism." http://www.nickdiakopoulos.com/2011/04/22/a-functional-roadmap-for-innovation-in-computational-journalism/.

Diakopoulos, Nicholas. 2015. "Algorithmic Accountability." *Digital Journalism* 3 (3):398–415. https://doi.org/10.1080/21670811.2014.976411.

Domingo, David, Pere Masip, and Irene Costera Meijer. 2015. "Tracing Digital News Networks." *Digital Journalism* 3 (1):53–67. https://doi.org/10.1080/21670811.2014.927996.

Donoghue, Andrew. 2008. "mySociety: Open Democracy, Open Source." http://www.h-online.com/open/features/mySociety-Open-democracy-open-source-746402.html.

Donohue, Stacy. 2016. "Engines of Change." *Civic Hall*. http://civichall.org/civicist/engines-of-change/.

Edwards, Gemma. 2010. "Mixed-Method Approaches to Social Network Analysis." Working Paper. ESRC NCRM. http://eprints.ncrm.ac.uk/842/.

Emmel, Nick. 2008. "Participatory Mapping: An Innovative Sociological Method." Working Paper. Real Life Methods. http://eprints.ncrm.ac.uk/540/.

Escher, Tobias. 2011. "Impact of UK Parliamentary Sites." Oxford Internet Institute. https://www.mysociety.org/research/impact-of-uk-parliamentary-sites/.

Ettema, James S., and Theodore Lewis Glasser. 1998. *Custodians of Conscience: Investigative Journalism and Public Virtue*. New York: Columbia University Press.

Fink, Katherine, and C. W. Anderson. 2015. "Data Journalism in the United States." *Journalism Studies* 16 (4):467–81. https://doi.org/10.1080/1461670X.2014.939852.

Flew, Terry, Christina Spurgeon, Anna Daniel, and Adam Swift. 2012. "The Promise of Computational Journalism." *Journalism Practice* 6 (2):157–71. https://doi.org/10.1080/17512786.2011.616655.

Garcelon, Marc. 2009. "An Information Commons? Creative Commons and Public Access to Cultural Creations." *New Media & Society* 11 (8):1307–26. https://doi.org/10.1177/1461444809343081.

Giridharadas, Anand. 2010. "Ushahidi - Africa's Gift to Silicon Valley: How to Track a Crisis." *The New York Times*, March. http://www.nytimes.com/2010/03/14/weekinreview/14giridharadas.html.

Gitelman, Lisa. 2013. *Raw Data Is an Oxymoron*. Cambridge, MA: MIT Press.

Glaser, Barney G, and Anselm L Strauss. 1967. *The Discovery of Grounded Theory: Strategies for Qualitative Research*. Chicago: Aldine Pub. Co.

Graves, Lucas. 2017. "The Monitorial Citizen in the 'Democratic Recession'." *Journalism Studies* 18 (10):1239–50. https://doi.org/10.1080/1461670X.2017.1338153.

Gray, Jonathan. 2016. "Datafication and Democracy: Recalibrating Digital Information Systems to Address Societal Interests." *IPPR*. https://www.ippr.org/juncture-item/datafication-and-democracy.

Gray, Jonathan, Liliana Bounegru, and Lucy Chambers, eds. 2012. *The Data Journalism Handbook*. 1st ed. Sebastopol, CA: O'Reilly Media. http://datajournalismhandbook.org/.

Gray, Jonathan, Danny Lämmerhirt, and Liliana Bounegru. 2016. "Changing What Counts: How Can Citizen-Generated and Civil Society Data Be Used

as an Advocacy Tool to Change Official Data Collection?" http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2742871.

Gregg, Melissa. 2015. "Hack for Good: Speculative Labour, App Development and the Burden of Austerity." *The Fibreculture Journal*, no. 25. http://twentyfive.fibreculturejournal.org/fcj-186-hack-for-good-speculative-labour-app-development-and-the-burden-of-austerity/.

Gurstein, Michael. 2011. "Open Data: Empowering the Empowered or Effective Data Use for Everyone?" *First Monday* 16 (2). https://doi.org/10.5210/fm.v16i2.3316.

Gynnild, Astrid. 2014. "Journalism Innovation Leads to Innovation Journalism: The Impact of Computational Exploration on Changing Mindsets." *Journalism* 15 (6):713–30. https://doi.org/10.1177/1464884913486393.

Hallberg, Lillemor R.-M. 2006. "The 'Core Category' of Grounded Theory: Making Constant Comparisons." *International Journal of Qualitative Studies on Health and Well-Being* 1 (3):141–48. https://doi.org/10.1080/17482620600858399.

Hamilton, James T., and Fred Turner. 2009. "Accountability Through Algorithm: Developing the Field of Computational Journalism." Report from Center for Advanced Study in the Behavioural Sciences, Summer Workshop. http://dewitt.sanford.duke.edu/wp-content/uploads/2011/12/About-3-Research-B-cj-1-finalreport.pdf.

Hanitzsch, Thomas, and Tim P. Vos. 2018. "Journalism Beyond Democracy: A New Look into Journalistic Roles in Political and Everyday Life." *Journalism* 19 (2):146–64. https://doi.org/10.1177/1464884916673386.

Hansen, Ejvind. 2012. "Aporias of Digital Journalism." *Journalism* 14 (5):678–94. https://doi.org/10.1177/1464884912453283.

Hepp, Andreas. 2016. "Pioneer Communities: Collective Actors in Deep Mediatisation." *Media, Culture & Society* 38 (6):918–33. https://doi.org/10.1177/0163443716664484.

Hepp, Andreas, Matthias Berg, and Cindy Roitsch. 2014. "Mediatized Worlds of Communitization: Young People as Localists, Centrists, Multi-Localists and Pluralists." In *Mediatized Worlds: Culture and Society in a Media Age*, edited by Andreas Hepp and Friedrich Krotz, 174–203. Palgrave.

Hermida, Alfred, and Mary Lynn Young. 2017. "Finding the Data Unicorn." *Digital Journalism* 5 (2):159–76. https://doi.org/10.1080/21670811.2016.1162663.

Holovaty, Adrian. 2006. "A Fundamental Way Newspaper Sites Need to Change." http://www.holovaty.com/writing/fundamental-change/.

Hope, Alexis. 2014. "Monitorial Citizenship: Projects and Tools." *MIT Center for Civic Media.* https://civic.mit.edu/blog/alexishope/monitorial-citizenship-projects-and-tools.

Howard, Alexander. 2014a. "Aron Pilhofer on Data Journalism, Culture and Going Digital." *The Tow Center for Digital Journalism.* http://towcenter.org/aron-pilhofer-on-data-journalism-culture-and-going-digital/.

Howard, Alexander. 2014b. "Debugging the Backlash to Data Journalism." *Two Center for Digital Journalism*. http://towcenter.org/debugging-the-backlash-to-data-journalism/.

Howard, Alexander. 2014c. "The Art and Science of Data-Driven Journalism." A Tow/Knight Report. Two Center for Digital Journalism. http://towcenter.org/wp-content/uploads/2014/05/Tow-Center-Data-Driven-Journalism.pdf.

IndigoTrust. 2011. "Tom Steinberg (mySociety) at Power of Information Conference." https://www.youtube.com/watch?v=3W6D8ZehqhU.

Janssen, Katleen. 2012. "Open Government Data and the Right to Information: Opportunities and Obstacles." *The Journal of Community Informatics* 8 (2). http://ci-journal.net/index.php/ciej/article/view/952.

Jenkins, Henry. 2006. *Convergence Culture: Where Old and New Media Collide*. NYU Press.

Johnson, Jeffrey Alan. 2014. "From Open Data to Information Justice." *Ethics and Information Technology* 16 (4):263–74. https://doi.org/10.1007/s10676-014-9351-8.

Josephi, Beate. 2012. "How Much Democracy Does Journalism Need?" *Journalism* 14 (4):474–89. https://doi.org/10.1177/1464884912464172.

Kapczynski, Amy. 2010. "Access to Knowledge: A Conceptual Genealogy." In *Access to Knowledge in the Age of Intellectual Property*, edited by Amy Kapczynski and Gaëlle Krikorian, 17–56. New York: Zone Books. http://www.zonebooks.org/titles/KRIK_ACC.html.

Karlsen, Joakim, and Eirik Stavelin. 2014. "Computational Journalism in Norwegian Newsrooms." *Journalism Practice* 8 (1):34–48. https://doi.org/10.1080/17512786.2013.813190.

Keane, John. 2009. *The Life and Death of Democracy*. 1st American ed. New York: W.W. Norton & Co.

Kelty, Christopher M. 2008. *Two Bits: The Cultural Significance of Free Software*. Experimental Futures. Durham: Duke University Press. http://twobits.net/download/index.html.

Kennedy, Helen. 2016. *Post, Mine, Repeat*. London: Palgrave Macmillan UK. http://link.springer.com/10.1057/978-1-137-35398-6.

Kennedy, Helen, and Rosemary Lucy Hill. 2017. "The Feeling of Numbers: Emotions in Everyday Engagements with Data and Their Visualisation." *Sociology*, February. https://doi.org/10.1177/0038038516674675.

Kennedy, Helen, and Giles Moss. 2015. "Known or Knowing Publics? Social Media Data Mining and the Question of Public Agency." *Big Data & Society* 2 (2). https://doi.org/10.1177/2053951715611145.

Kennedy, Helen, Thomas Poell, and Jose van Dijck. 2015. "Data and Agency." *Big Data & Society* 2 (2). https://doi.org/10.1177/2053951715621569.

Kitchin, R. 2014. "Big Data, New Epistemologies and Paradigm Shifts." *Big Data & Society* 1 (1). https://doi.org/10.1177/2053951714528481.

Knight Foundation. 2013. "Trends in Civic Tech." http://www.knightfoundation.org/features/civictech/.

Konieczna, Magda, and Elia Powers. 2017. "What Can Nonprofit Journalists Actually Do for Democracy?" *Journalism Studies* 18 (12):1542–58. https://doi.org/10.1080/1461670X.2015.1134273.

Krotoski, Aleks. 2010. "Interview: Tom Steinberg (MySociety)." *Untangling the Web*. http://untanglingtheweb.tumblr.com/post/1650115558/interview-tom-steinberg-mysociety.

Krotz, Friedrich. 2005. *Neue Theorien entwickeln: eine Einführung in die Grounded Theory, die Heuristische Sozialforschung und die Ethnographie anhand von Beispielen aus der Kommunikationsforschung*. Von Halem.

Kvale, Steinar. 1996. *InterViews. An Introduction to Qualitative Research Interviewing*. Thousand Oaks: Sage.

Larson, Jeff, Surya Mattu, Lauren Kirchner, and Julia Angwin. 2016. "How We Analyzed the COMPAS Recidivism Algorithm." *ProPublica*. https://www.propublica.org/article/how-we-analyzed-the-compas-recidivism-algorithm.

Latour, Bruno. 2005. *Reassembling the Social: An Introduction to Actor-Network-Theory*. Clarendon Lectures in Management Studies. Oxford; New York: Oxford University Press.

Lesage, Frédérik, and Robert A Hackett. 2014. "Between Objectivity and Openness - the Mediality of Data for Journalism." *Media and Communication* 1 (1):39–50. https://doi.org/10.12924/mac2013.01010039.

Levitas, Jake. 2013. "Defining Civic Hacking. How a Common Framework Can Unite New Forms of Engagement." *Code for America Blog*. http://www.codeforamerica.org/2013/06/07/defining-civic-hacking/.

Levy, Steven. 1984. *Hackers: Heroes of the Computer Revolution*. 1st ed. Garden City, N.Y: Anchor Press/Doubleday.

Lewis, Seth C. 2012a. "From Journalism to Information: The Transformation of the Knight Foundation and News Innovation." *Mass Communication and Society* 15 (3):309–34. https://doi.org/10.1080/15205436.2011.611607.

Lewis, Seth C. 2012b. "The Tension Between Professional Control and Open Participation. Journalism and Its Boundaries." *Information, Communication & Society* 15 (6):836–66. https://doi.org/10.1080/1369118X.2012.674150.

Lewis, Seth C., and Nikki Usher. 2013. "Open Source and Journalism: Toward New Frameworks for Imagining News Innovation." *Media Culture Society* 35 (5):602–19. https://doi.org/10.1177/0163443713485494.

Lewis, Seth C., and Nikki Usher. 2014. "Code, Collaboration, and the Future of Journalism." *Digital Journalism* 2 (3):383–93. https://doi.org/10.1080/21670811.2014.895504.

Lewis, Seth C., and Nikki Usher. 2016. "Trading Zones, Boundary Objects, and the Pursuit of News Innovation A Case Study of Journalists and Programmers." *Convergence: The International Journal of Research into New Media Technologies* 22 (5):543–60. https://doi.org/10.1177/1354856515623865.

Lievrouw, Leah A. 2011. *Alternative and Activist New Media*. Digital Media and Society Series. Cambridge, UK ; Malden, MA: Polity.

Loosen, Wiebke, Julius Reimer, and Fenja Schmidt. 2015. "When Data Become News A Content Analysis of Data Journalism Pieces." In. Cardiff University, UK.

Lorenz, Mirko. 2010. "Data-Driven Journalism: What Is There to Learn?" In. University of Amsterdam. http://mediapusher.eu/datadrivenjournalism/pdf/ddj_paper_final.pdf.

Lorenz, Mirko, Nicolas Kayser-Bril, and Geoff McGhee. 2011. "Media Companies Must Become Trusted Data Hubs." *DataDrivenJournalism.net*. http://datadrivenjournalism.net/news_and_analysis/media_companies_must_become_trusted_data_hubs.

Maeyer, Juliette De, Manon Libert, David Domingo, François Heinderyckx, and Florence Le Cam. 2015. "Waiting for Data Journalism." *Digital Journalism* 3 (3):432–46. https://doi.org/10.1080/21670811.2014.976415.

Mansell, Robin. 2012. *Imagining the Internet: Communication, Innovation, and Governance*. Oxford: Oxford University Press.

Margetts, Helen. 2013. "Data, Data Everywhere: Open Data Versus Big Data in the Quest for Transparency." In *Transparency in Politics and the Media: Accountability and Open Government*, edited by Nigel Bowles, James T. Hamilton, and David A. L. Levy, 167–78. London: I.B.Tauris.

Matei, Ani, and Sergiu Ioan Irimia. 2014. "Open Source Governance. A More Ambitious Cousin of Collaborative Governance." *International Journal of Public Administration* 37 (12):812–23. https://doi.org/10.1080/01900692.2014.907315.

Mayer-Schönberger, Viktor, and Kenneth Cukier. 2013. *Big Data: A Revolution That Will Transform How We Live, Work and Think*. London: Murray.

McGregor, Susan. 2013. "CAR Hits the Mainstream." *Columbia Journalism Review*. http://www.cjr.org/data_points/computer_assisted_reporting.php.

McVeigh-Schultz, Joshua, and Nancy K. Baym. 2015. "Thinking of You: Vernacular Affordance in the Context of the Microsocial Relationship App, Couple." *Social Media + Society* 1 (2). https://doi.org/10.1177/2056305115604649.

MediaStandardsTrust. 2009. "Why Journalism Matters: Alan Rusbridger (Part 3)." https://www.youtube.com/watch?time_continue=19\&v=F5FazLeT430.

Meyer, Philip. 2002. *Precision Journalism: A Reporter's Introduction to Social Science Methods*. Lanham, Md.: Rowman & Littlefield Publishers.

Milan, Stefania, and Lonneke Van Der Velden. 2016. "The Alternative Epistemologies of Data Activism." SSRN Scholarly Paper ID 2850470. Rochester, NY: Social Science Research Network. https://papers.ssrn.com/abstract=2850470.

mySociety. 2011. "You Need Volunteers to Make Your Website Work." *mySociety*. https://www.mysociety.org/2011/07/29/you-need-volunteers-to-make-your-website-work/.

mySociety. 2014. "Find Out About SayIt." https://www.youtube.com/watch?v=x7wHMv_gQ2w.

mySociety. 2015. "mySociety Year in Review 2015." https://2015.mysociety.org/.

mySociety. 2016. "SayIt by mySociety: Online Meeting Transcripts." https://www.youtube.com/watch?v=Hzhcz7ic7Do.

mySociety. n.d. "mySociety." *mySociety Homepage*. Accessed July 14, 2016a. https://www.mysociety.org/.

mySociety. n.d. "Questions About SayIt." *SayIt Homepage*. Accessed August 16, 2016b. http://sayit.mysociety.org/about/questions.

Nafus, Dawn, and Jamie Sherman. 2014. "This One Does Not Go up to 11: The Quantified Self Movement as an Alternative Big Data Practice." *International Journal of Communication* 8:1784–94. http://ijoc.org/index.php/ijoc/article/view/2170.

Nagy, Peter, and Gina Neff. 2015. "Imagined Affordance: Reconstructing a Keyword for Communication Theory." *Social Media + Society* 1 (2). https://doi.org/10.1177/2056305115603385.

Nestoria. 2008. "Nestoria Interview - Tom Steinberg - mySociety." http://blog.nestoria.co.uk/post/44316704332/nestoria-interview-tom-steinberg-mysociety.

Nussbaum, Emily. 2009. "The New Journalism: Goosing the Gray Lady." *NYMag.com*. http://nymag.com/news/features/all-new/53344/.

OpenGovData.org. 2007. "The 8 Principles of Open Government Data." https://opengovdata.org/.

Open Knowledge. 2012. *The Open Data Handbook*. opendatahandbook.org.

Open Knowledge. n.d. "The Open Definition." Accessed January 30, 2015a. http://opendefinition.org/.

Open Knowledge. n.d. "What Is Open?" Accessed February 11, 2015b. https://okfn.org/opendata/.

OpenNews. n.d. "The Knight-Mozilla Fellowships." Accessed May 8, 2015. http://opennews.org/what/fellowships/.

O'Reilly, Tim. 2010. "Government as a Platform." In *Open Government*, edited by Daniel Lathrop and Laurel Ruma. O'Reilly Media. http://chimera.labs.oreilly.com/books/1234000000774/ch02.html.

Parasie, Sylvain. 2011. "'Hacker' Journalism - A New Utopia for the Press?" *Books & Ideas*. http://www.booksandideas.net/Hacker-Journalism-A-New-Utopia-for.html.

Parasie, Sylvain, and Eric Dagiral. 2013. "Data-Driven Journalism and the Public Good: "Computer-Assisted-Reporters" and "Programmer-Journalists" in Chicago." *New Media & Society* 15 (6):853–71. https://doi.org/10.1177/1461444812463345.

Pariser, Eli. 2011. *The Filter Bubble: What the Internet Is Hiding from You*. 1ST edition. New York: Penguin Press HC, The.

Petre, Caitlin. 2013. "A Quantitative Turn in Journalism?" *Tow Center for Digital Journalism*. http://towcenter.org/blog/a-quantitative-turn-in-journalism/.

Polletta, Francesca. 1999. "'Free Spaces' in Collective Action." *Theory and Society* 28 (1):1–38. https://doi.org/10.1023/A:1006941408302.

Raymond, Eric S. 2001. *The Cathedral & the Bazaar: Musings on Linux and Open Source by an Accidental Revolutionary*. Cambridge, Mass.: O'Reilly.

Reich, Zvi. 2013. "The Impact of Technology on News Reporting: A Longitudinal Perspective." *Journalism & Mass Communication Quarterly* 90 (3):417–34. https://doi.org/10.1177/1077699013493789.

Rogers, Richard. 2013. *Digital Methods*. Cambridge, Massachusetts: The MIT Press.

Rosanvallon, Pierre. 2008. *Counter-Democracy: Politics in an Age of Distrust*. John Robert Seeley Lectures 7. Cambridge, UK ; New York: Cambridge University Press.

Rose, Nikolas. 1999. *Powers of Freedom: Reframing Political Thought*. Port Chester, NY, USA: Cambridge University Press.

Royal, Cindy. 2010. "The Journalist as Programmer: A Case Study of the New York Times Interactive News Technology Department." In *The International Symposium in Online Journalism*. The University of Texas.

Sauter, Molly. 2014. *The Coming Swarm: DDOS Actions, Hacktivism, and Civil Disobedience on the Internet*. New York: Bloomsbury Academic.

Schatzki, Theodore R. 2001. "Introduction: Practice Research." In *The Practice Turn in Contemporary Theory*, edited by Theodore R. Schatzki, Karin Knorr-Cetina, and Eike von Savigny, 10–23. Ebrary Ebrary Academic Complete. New York: Routledge.

Schatzki, Theodore R. 2012. "A Primer on Practices." In *Practice-Based Education: Perspectives and Strategies*, edited by Joy Higgs, Ronald Barnett, and Stephen Billett, 13–26. Rotterdam: Sense Publishers.

Schrock, Andrew. 2016. "Civic Hacking as Data Activism and Advocacy: A History from Publicity to Open Government Data." *New Media & Society* 18 (4):581–99. https://doi.org/10.1177/1461444816629469.

Schudson, Michael. 1998. *The Good Citizen: A History of American Civic Life*. New York: Martin Kessler Books.

Schudson, Michael. 2008. "The "Lippmann-Dewey Debate" and the Invention of Walter Lippmann as an Anti-Democrat 1985-1996." *International Journal of Communication* 2:12. http://ijoc.org/index.php/ijoc/article/view/343.

Schudson, Michael. 2015. *The Rise of the Right to Know: Politics and the Culture of Transparency, 1945-1975*. Harvard University Press.

Schwegmann, Claudia. 2012. "It's a Fad! 7 Reasons Against Open Data." *Open Knowledge Foundation Deutschland*. http://okfn.de/blog/2012/11/7-reasons-against-open-data/.

Shaw, Emily. 2016. "Debugging Democracy: What Is It We're Doing Here, People?, Agreeing on the Problems, Agreeing on the Solutions, Agreeing to See If It's Working." *Medium*. https://medium.com/@emilydshaw/debugging-democracy-bfa68e37967b\#.tm7gua558.

Shove, Elizabeth. 2003. *Comfort, Cleanliness and Convenience: The Social Organization of Normality*. New Technologies/New Cultures. Oxford: Berg.

Sifry, Micah L. 2014. "Civic Tech and Engagement: In Search of a Common Language." *TechPresident*. http://techpresident.com/news/25261/civic-tech-and-engagement-search-common-language.

Singer, Jane B. 2005. "The Political J-Blogger: 'Normalizing' a New Media Form to Fit Old Norms and Practices." *Journalism* 6 (2):173–98. https://doi.org/10.1177/1464884905051009.

Slee, Tom. 2012. "Seeing Like a Geek." *Crooked Timber*. http://crookedtimber.org/2012/06/25/seeing-like-a-geek/.

Steinberg, Tom. 2012. "mySociety's Components Strategy Take on Small Pieces Loosely Joined." *mySociety*. https://www.mysociety.org/2012/07/04/mysocietys-components-strategy-our-take-on-small-pieces-loosely-joined/.

Steinberg, Tom. 2013. "What Should We Do About the Naming Deficit/Surplus?" *mySociety*. https://www.mysociety.org/2013/04/09/what-should-we-do-about-the-naming-deficitsurplus/.

Stempeck, Matt. 2016. "Towards a Taxonomy of Civic Technology." *Microsoft on the Issues*. https://blogs.microsoft.com/on-the-issues/2016/04/27/towards-taxonomy-civic-technology/.

Strauss, Anselm L., and Juliet M. Corbin. 1998. *Basics of Qualitative Research: Techniques and Procedures for Developing Grounded Theory*. 2nd ed. Thousand Oaks: Sage Publications.

Sunlight Foundation. 2010. "Ten Principles for Opening up Government Information." http://sunlightfoundation.com/policy/documents/ten-open-data-principles/.

Taylor, Charles. 2004. *Modern Social Imaginaries*. Public Planet Books. Durham: Duke University Press.

Tkacz, Nathaniel. 2012. "From Open Source to Open Government: A Critique of Open Politics." *Ephemera: Theory and Politics in Organization* 12 (4):386–405. http://wrap.warwick.ac.uk/53295.

Tkacz, Nathaniel. 2013. "Open Sesame." *Aeon Magazine*. http://aeon.co/magazine/society/nathaniel-tkacz-open-source-government/.

Tkacz, Nathaniel. 2015. *Wikipedia and the Politics of Openness*. Chicago; London: University of Chicago Press.

Townend, Judith. 2008. "mySociety Turns Five - How Can Journalists Best Use Its Sites?" *Journalism.co.uk*. https://www.journalism.co.uk/news-features/mysociety-turns-five--how-can-journalists-best-use-its-sites-/s5/a532502/.

Townend, Judith. 2009. "The Growth of Online Watchdogs: Are They 'Journalism' and Does It Matter?" https://blogs.journalism.co.uk/2009/07/24/the-growth-of-online-watchdogs-are-they-journalism-and-does-it-matter/.

Tuomi, Ilkka. 1999. "Data Is More Than Knowledge: Implications of the Reversed Knowledge Hierarchy for Knowledge Management and Organizational Memory." *Journal of Management Information Systems* 16 (3):103–17. http://www.jstor.org/stable/40398446.

UsNowFilm. 2008. "Tom Steinberg - Personal Interest, Public Value." https://www.youtube.com/watch?v=4L2kquBCpnM.

van Dijck, José. 2014. "Datafication, Dataism and Dataveillance: Big Data Between Scientific Paradigm and Ideology." *Surveillance & Society* 12 (2):197–208. http://library.queensu.ca/ojs/index.php/surveillance-and-society/article/view/datafication.

van Dijck, José, and Thomas Poell. 2013. "Understanding Social Media Logic." *Media and Communication* 1 (1):2–14. https://doi.org/10.12924/mac2013.01010002.

Wahl-Jorgensen, Karin. 2014. "Is WikiLeaks Challenging the Paradigm of Journalism? Boundary Work and Beyond." *International Journal of Communication* 8:2581–92. http://ijoc.org/index.php/ijoc/article/view/2771.

Weber, Steve. 2004. *The Success of Open Source*. Cambridge, MA: Harvard University Press.

Weinacht, Stefan, and Ralf Spiller. 2014. "Datenjournalismus in Deutschland." *Publizistik* 59 (4):411–33. https://doi.org/10.1007/s11616-014-0213-5.

Weinstein, Matthew. 2006. "TAMS Analyzer. Anthropology as Cultural Critique in a Digital Age." *Social Science Computer Review* 24 (1):68–77. https://doi.org/10.1177/0894439305281496.

Weiss, Robert Stuart. 1994. *Learning from Strangers: The Art and Method of Qualitative Interview Studies*. New York: Free Press.

Wenger, Etienne. 1998. "Communities of Practice: Learning as a Social System." *Systems Thinker* 9 (5):2–3. https://moo27pilot.ied.edu.hk/pluginfile.php/415222/mod_resource/content/3/Learningasasocialsystem.pdf.

Wing, Jeannette M. 2006. "Computational Thinking." *Commun. ACM* 49 (3):33–35. https://doi.org/10.1145/1118178.1118215.

Wing, Jeannette M. 2010. "Computational Thinking: What and Why?" Pittsburgh. https://www.cs.cmu.edu/~CompThink/papers/TheLinkWing.pdf.

www.parliament.uk. 2014. "Review of Parliament's Online Services Report Published." http://www.parliament.uk/business/news/2014/march/review-of-parliaments-online-services-report/.

Yin, Robert Kuo-zuir. 2014. *Case Study Research: Design and Methods*. 5th ed. Thousand Oaks, CA etc.: Sage Publications.

Yu, Harlan, and David G. Robinson. 2012. "The New Ambiguity of 'Open Government'." *SSRN Electronic Journal*, February. https://doi.org/10.2139/ssrn.2012489.

Zuboff, Shoshana. 2015. "Big Other: Surveillance Capitalism and the Prospects of an Information Civilization." *Journal of Information Technology* 30 (1):75–89. https://doi.org/10.1057/jit.2015.5.

Zuckerman, Ethan. 2014a. "Promise Tracker and Monitorial Citizenship."...*My Heart's in Accra*. http://www.ethanzuckerman.com/blog/2014/01/24/promise-tracker-and-monitorial-citizenship/.

Zuckerman, Ethan. 2014b. "New Media, New Civics?" *Policy & Internet* 6 (2):151–68. https://doi.org/10.1002/1944-2866.POI360

Appendix A: The process of data collection and analysis

In the following, I describe the process of data collection and analysis for each case study in more detail (for a description of the general steps in constructivist grounded theory, see Chapter 3). Note that in grounded theory, the collection and analysis of data overlap due to theoretical sampling. In practice, the transitions from initial to focused coding, as well as the development of the theories overall, were very iterative, constantly moving from the data to new theoretical ideas. For better comprehension, however, I present the different spirals through the process of grounded theory as clearly separated (for an explanation of 'spiral', see Chapter 3).

First case study

I decided to conduct a case study about the Open Knowledge Foundation Germany (OKF DE) because it is one of the most visible actors in the German open data movement, and it has created several nationally well-known and influential civic tech applications (e.g. the freedom of information website FragDenStaat, or the parliamentary monitoring website OffenesParlament). I first contacted the organization through its official mailing list, where I introduced myself and my research project. This introduction led to the first two interviews in September 2012. The first interview was a group interview with three members (a senior developer and two members of the project management team) in Berlin; the second interview took place in Bremen with a committee member. For this initial data sample, the interview guide included a couple of broad questions in order to get more familiar with the subject and

the organization. I also included central self-portraying documents in the analysis: the official homepage of the OKF DE, the 'Open Definition' (Open Knowledge n.d.) and the 'Open Data Handbook' (Open Knowledge 2012).

This formed my initial data sample. In this first case study, I was primarily interested in the social imaginaries of activists and the role of (open) data in those imaginaries. I initially coded for practices, goals and self-understanding. Practices were signaled by verbs and often overlapped with codes that described goals. Following Schatzki's (2012, 2001) understanding of practices described in Chapter 1, the goals captured the purposes of the practices I identified. If possible, I connected practices and goals to specific projects. The parliamentary monitoring website OffenesParlament, for example, was connected to goals like *more transparency* and *more engagement*, as well as to practices such as *structuring data*.[28] After openly coding the data this way, I explored the relationship between the different concepts in focused coding, where I tried to merge similar codes into more abstract concepts. After this merging and abstracting was complete, I went through all the segments of data that were summarized under each category to more clearly identify their characteristics, which subsequently sparked ideas to refine or to combine them. Throughout this process, I made intensive use of mind mapping software and memo-writing to explore the dimensions and relationships between categories.

Based on the analysis of the initial data sampling, I found that the self-understanding as 'civic hackers' is very useful to address my research question.

[28] In this chapter, italicized words are used for the codes or categories that I developed during my analysis.

Therefore, I participated in a workshop organized by the OKF DE in November 2012 called 'stadtlandcode' (which was essentially a prototype for the 'Code for Germany' initiative, see Chapter 6). This workshop was the first attempt of the OKF DE to build a civic tech community in Germany. Members of the OKF DE translated the term 'civic tech' into German for the first time and explained its meaning to both participants and funders. In addition, the workshop was organized as a contest in which participants should pitch ideas for civic tech applications. The way the OKF DE attempted to shape the meaning of the term in the German context, and how the civic tech projects pitched by the participants of the workshop were discussed and evaluated generated rich data on the goals, self-understandings and practices of the OKF DE members.

During the event, I acted as a participant observer writing down notes from presentations, impressions, questions. Those notes informed four new interviews during the workshop (two more project managers, two developers). Another interview with the chairman and founder was conducted the day after the workshop. This data formed the basis for the second spiral of my study. During the analysis of the data, I also contacted two OKF DE members I interviewed previously to ask specific follow-up questions. This led to new data from email responses and a Skype interview.

Based on a theoretical sampling, I also decided to contact another member of the OKF DE because it had a very different background and education compared to the other members. I was interested in exploring if I could still identify similar practices, ambitions and self-understandings. I contacted this member and interviewed it in January 2013 in Hannover. For the data analysis in this second spiral, I repeated initial coding on the new data and compared it with the old one. I also applied my previous categories to the new data, with

the result that some of these categories were modified or dismissed, others became more central. In a third and final spiral, I included new online content following a theoretical sampling that helped to further refine and elaborate the theory: interviews members of the OKF DE gave elsewhere, articles they wrote about their work, or about open data in general.

Throughout these different spirals and refinements, I eventually developed a core category that integrates and elucidates the central aspects of my theory: *spreading open knowledge*. Spreading open knowledge means that members of the OKF DE aim to diffuse raw (government) data that is accessible through applications (like interactive visualizations). For the OKF DE, data forms the basis for generating knowledge. In other words, data becomes knowledge when it is interpreted, meaningful, and actionable. All other practices identified in the analysis are subcategories of this: *constructing open infrastructures*, *creating and changing empowering intermediaries*, *defining openness*, and *lobbying*. The goal of spreading open knowledge is to *empower citizens*, which means *more opportunities for participation*, a more *fact-oriented public discourse*, and *improved accountability*. The self-understanding of the OKF DE was described as a *civic developer network*.

The data collection and the development of the category system were conducted in relation to my MA thesis, which formed the basis for my PhD project proposal. As it was still relevant for my PhD, I further elaborated the analysis with an extensive literature review, moving from a rather descriptive study to a theory about the social imaginaries of open data activists and the implications for the agency of publics. One of the most important sources for evolving the theory was Kelty's (2008) study about the cultural significance of free software (see Chapter 2). Correlating my categories with the core 'components' of free software identified by Kelty (2008), I was able to show that the open data movement is a modulation of open source. Members of the

OKF DE apply practices and values from open source culture to the creation and use of data. See Chapter 4 for more details.

Second case study

The first contact with mySociety was established at the stadtlandcode workshop mentioned in the previous section. At this workshop, Tony Bowden, a long-standing member of mySociety, was invited as a speaker to give advice on how to create successful civic tech projects. He used several examples from mySociety in the UK, explained how they work and what makes them successful. I made extensive notes of this presentation. In subsequent interviews with members of the OKF DE, I found that mySociety's work has a huge influence and represents an international 'best practice' of civic tech. The second encounter was at the Open Knowledge Festival 2014, which was attended by several mySociety members. The field notes collected at these encounters were expanded with online content: mySociety's homepage (including several blog posts) and websites of key projects (WhatDoTheyKnow, FixMyStreet, TheyWorkForYou, WriteToThem). I also included older version of mySociety's main homepage using the Wayback Machine of archive.org. This data informed the guides for the first two interviews, which were conducted in May 2015 with a senior developer and a member of mySociety's international team.

This data formed my initial sampling and was the basis for the initial coding. Due to my interest in how mySociety uses data, I began with coding both verb-by-verb and project-by-project. For example, interviewees used verbs such as *helping*, *annotating*, *deep linking*, *pushing*, *scraping* when they spoke about mySociety's parliamentary monitoring website TheyWorkForYou. Following Charmaz's (2006, 49) advice to code for actions with gerunds to avoid applying

pre-conceived concepts to the data (e.g. by coding with 'stating' rather than 'statement'), I also developed some slightly more abstract codes that did not relate to specific projects, but to mySociety's broader mission and self-understanding: *empowering*, *community building*, *increasing reusability* and others. After I coded all the material this way, I combined similar verbs into more abstract practices. During focused coding, the developed categories were tested across all the collected materials to examine their validity, and to create a hierarchy of categories and subcategories.

As with the other case studies, I made intensive use of mind mapping software and memo-writing to develop theoretical sensibility. I first explored the relationship between categories and determined, for example, that the categories *deep linking into documents* and *emphasizing usability* are both subcategories of *improving accessibility*; while *improving accessibility* itself is a subcategory of *developing empowering services*. Again, I frequently reviewed the data that was summarized under each category to further elaborate the codes and build levels of abstraction. *Developing empowering services*, for example, entails codes like *engaging citizens*, *making advocacy more effective*, *complementing public institutions*, or *emphasizing user experience*. Using memos, I brainstormed about the meaning of *developing empowering services* in each of these codes, and explored other potentially relevant aspects that were not covered in the initial data sampling.

Following my initial sampling, I conducted a first theoretical sampling. Another in-depth interview with an updated interview guide was conducted with another senior developer in June 2016. This developer was responsible for mySociety's international freedom of information platform Alavateli. mySociety develops customized versions of Alavateli in more than 20 countries. Interviewing this member promised to give more insights into how

the tools of mySociety reflect certain values and ambitions, and the role of international work for mySociety. Moreover, I included more online content in the analysis that helped to address open questions in the evolving theory (e.g. blog posts from mySociety's homepage that helped exploring my categories). A third and final spiral included two additional interviews in August 2015 (with the founder and former CEO, as well as a member of mySociety's research team), and additional content taken from discussion forums and participant observations at the Mozilla Festival in November 2015.

Similar to the first case study, I eventually developed a core category: *facilitating engagement*, which is a more refined version of the *developing empowering services* category I used in the analysis of the initial data sample. In short, mySociety understands itself as a facilitator of civic engagement and imagines that data – and more specifically the affordances of structured data – can be used to facilitate engagement by *improving government accessibility*, *influencing governments*, *demonstrating impact* to users and *scaling technical solutions*:

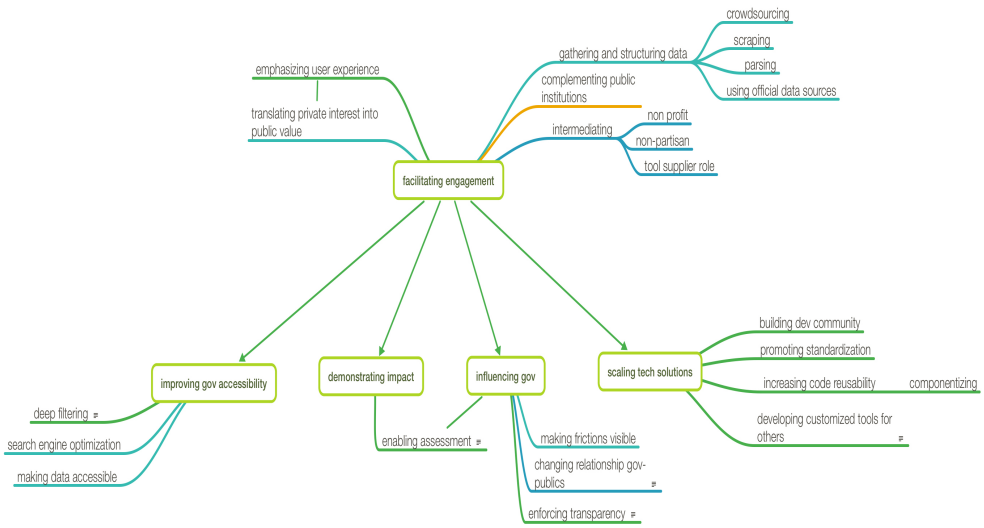


Figure 3: Overview of the final category system for the mySociety case study

Through facilitating engagement in these various ways, mySociety supports a form of participatory culture. The ways in which data is used by mySociety is intrinsically tied to this broader mission and therefore all data-related practices are connected to it. For a more detailed explanation, see Chapter 5.

Third case study

In contrast to the first two case studies, the collection and analysis of data for the third study on the entanglements between data journalists and data activists did not focus on any specific organization and relied much more on participant observations and interviews. In February 2016, I moved to Berlin and was able to immerse myself in the local data journalism and civic tech scene. Before I moved, I already established contacts with many data journalists across Germany at the Daten Labor conference in Dortmund (October 2015). Throughout my stay in Berlin, I regularly joined local meetups organized via meetup.com (Hacks/Hackers, #DDJ Berlin, Code for Germany and others). In total, I conducted 14 interviews between February and September 2016 with data journalists working in different national or local news media with different organizational structures (traditional news media organizations, startups, or non-profit newsrooms). While the different spirals of data collection and analysis in the other case studies were relatively separate, the process of analysis and data collection guided by theoretical samplings was much more iterative and continuous due to my direct local involvement.

As an example, during a #DDJ Berlin ('Data-Driven Journalism') meetup in February 2016, a journalist working at a national newspaper presented a project about the voting behavior of German politicians. This project was interesting for me because it was explicitly inspired by civic tech applications, some of which were developed by mySociety. I contacted this journalist

afterwards to set up an interview where I further explored this connection using my notes from the meetup and additional content, taken from the personal homepage and other articles written by the interviewee. At the end of each interview, I asked to whom I should talk next. This gave me new ideas for my data collection and provided some insights into who my interviewees deemed particularly relevant and/or representative.

Initially, the selection of interviewees was analogous to the OKF DE and mySociety case studies. I was interested in collecting the 'best practice' of data journalism in Germany by interviewing data journalists from leading German news media. At a relatively early stage of the data analysis it became clear, however, that there were significant differences in terms of practices, self-understandings and views on the relationship between data journalism and civic tech that could not be integrated into one coherent category system. Grouping interviewees with compatible and overlapping practices and views (I initially coded verb-by-verb again and gradually created more abstract categories) revealed the importance of professional and organizational backgrounds. For example, data journalists working in traditional national news media are very similar to each other, but together they differed significantly from data journalists working in startups. Often, these differences were already visible by the frequency of certain codes. As an example, I coded if and how the interviewees identified with the title 'data journalist', and what it means to them. For most journalists working in traditional national news media, the term was unproblematic. They understood data journalism as a *method* of *investigative journalism*. By contrast, the majority of interviewees working in startups or other nontraditional settings with a more technological background *problematized* the term and *renegotiated* the meaning of journalism.

Following the logic of theoretical sampling, I subsequently tried to systematically interview data journalists in different organizational settings to verify and further explore these groups. I tried to find more interviewees working in startups and explored differences within each group. For example, there was a chance that the views of data journalists working in public broadcasting differ in significant ways from those working in commercial daily news media, so I contacted and arranged an interview with a journalist working at BR Data, the data journalism team of the public broadcaster Bayrischer Rundfunk. To be clear, this study did not try to gather a representative sample of data journalists working in Germany, but rather to achieve theoretical sufficiency for the grounded theory it developed (see Chapter 3 for an evaluation).

After I sharpened the profile of each groups, I returned to the data of the OKF DE and mySociety case studies and looked for similarities and overlaps. How did data activists relate to data journalism, and are there overlaps in self-understanding and practices between data activists and the different groups of data journalists I identified? The participatory maps collected during the interviews also helped to examine how data journalists and data activists form communities (see Chapter 3). I coded each map similar to the interviews and noted what types of groups or organizations were mentioned. Looking at the data this way, I realized that there is no clear separation between the different groups, but different ways of emphasizing and rationalizing a shared continuum of practices. At the one end of the spectrum is the core category identified in the case study of mySociety: *facilitating*, i.e. enabling others to take action themselves. At the other end of the spectrum was a new, second core category: *gatekeeping* in the traditional sense of journalists acting as gatekeepers of publicly relevant information. The different groups identified in this study

can be defined and delineated along these axes. Three categories helped to describe and to profile each group:

1. *Positioning*: How are the different groups positioning themselves within the public arena? Journalists working in traditional national news media created a clear distinction between the inside and outside of journalism, while those working in startups displayed a much stronger influence of the technology sector.
2. *Self-understanding*: How do different groups understand their own work? This category is an extension of positioning, but it does not fully overlap. Positioning is about relating one's own work to others, while self-understanding is about positively defining it.
3. *Data usage*: This category captures how different groups understand and use data. Two main subcategories are used: gatekeeping correlates with *story-driven* uses of data, while facilitating correlates with an emphasis on *granularity and completeness*. These subcategories are inspired by Parasie and Dagiral's (2013) study of computer-assisted reporters and programmer-journalists in Chicago.

Collectively, the categories and the relationships between them demonstrate how the different groups 'intermeshed' in the figurational sense outlined by Couldry and Hepp (2017): the different practices of each group are not necessarily competing or excluding one another, but are rather complementing and in some cases mutually reinforcing each other. What brings data journalists and data activists together, to adopt the vocabulary of the figurational approach, is the way in which different practices are interlocking. I also coded the different ways in which this interlocking takes place: *providing infrastructure*,

unlocking opportunities, *providing ideas and showing needs* etc. See Chapter 6 for more details.

Appendix B: Exemplary interview guide

The following is the generic interview guide used for the second case study about mySociety (Chapter 5). For each interviewee, it was modified and expanded to ask more specific questions about her or his work, or about statements the interviewees made elsewhere.

1. Opening: Involvement at mySociety and personal development

Opening question:

- Can you tell me how you got involved in mySociety?
  - When was this?
  - What led you to work for mySociety?

This opening was supposed to give an impression of what the interviewee is doing at mySociety, and provide insight into her or his personal motivations and aims. I followed with questions about how the work at mySociety evolved over time to get an impression of the range of activities:

- Since you started, has your job changed?
- Can you describe your work for mySociety today? Maybe you could walk me through a typical working day?

2. Exploration of a particular project

Each interviewee was asked to describe a particular project in-depth: what was the original purpose of this project, how was it designed to fulfill that purpose, and what role did data play in it. I opened with asking, for example:

- What projects are you mainly working on today?
- Were you involved in project X from the start?
- Can you tell me how it all started for you, the first time you were confronted with the idea and how you got involved in the project?
  - What was the original purpose and aim of this project, what problem was it meant to solve in your eyes?
- Alternatively: Was there a major change in the project since you were involved? Can you tell me more about this?

With these questions, I developed a deeper understanding of the goals and aims of the project. Next, I transitioned into how the project was implemented:

- Could you walk me through the development process that followed, how these original ideas were put into practice?
  - Can you remember a meeting or discussion that was key for you, that best illustrates the approach you took?

If it was not mentioned already by the interviewee, I would ask about the role of data within this specific project:

- Was, or is, gathering and/or analyzing data important for the project?
  - Can you explain what kind of data you collect, how you collect it, and how it is used?

3. Contrasting example(s)/generalization

After one project was explored in-depth, I explored how representative this project is for the work of my interviewees. Opening questions were:

- When you 'zoom out' and think about all the different projects you are involved in, would you say that project X is a good example to illustrate your work as a whole?
  - Are the purposes and aims of other projects you are working on similar to project X?/Are there different ways to archive these purposes?
  - Is there any project that you think is very different? If so, can you tell me more about it and how it is different for you?

I also asked more specifically about the role of data again:

- Beyond this specific example, would you say that there are certain categories of data that you are interested in, or certain types of data that are particularly interesting for mySociety in general?

Because 'structured' and 'machine-readable' data was so prominent on mySociety's websites, I often asked:

- When I look at the descriptions online, a term that is often used is 'structured' or 'machine-readable' data. Could you explain what it means and why it is important?
  - Maybe in relation to a specific project?

4. Self-understanding and motivations

In this part of the interview, I invited my interviewees to reflect on their self-understanding and personal ambitions:

- Would you say that there is something that you personally want to accomplish with your work?

- If you would have to give yourself a job title that would best describe what you are doing, what would it be?

If it was not mentioned, I asked specifically about *civic hacking* or *civic tech*:

- On the website, mySociety describes itself as a 'civic tech organization'. Do you consider yourself as a civic hacker – why (not)?
- What is civic hacking and civic tech about for you?
  - Is there a way for you to identify civic tech applications?

After encouraging interviewees to generalize their personal motivations and practices, I asked about their views on mySociety as an organization:

- We have now talked about the aims of the projects you are working on, and what you want to accomplish with your work. I would like to ask you to 'zoom out' again and think about mySociety as a whole: What do you see as the main mission of mySociety?
  - Is this a shared view, so do other members of mySociety view it in this way as well?
  - And is this the formal mission of the organization, of how the organization presents itself?

In connection to these questions about the overall mission, I asked:

- Do you think mySociety serves a public interest?
  - If so, how?

Asking explicitly about serving a public interest was a helpful preparation for the next set of questions about the relation to other organizations. Moreover, I was interested in comparing how my interviewees describe the way they serve the public with traditional journalistic values.

5. Participatory mapping: Relation to other organizations

How does my interviewee understand the larger network mySociety is involved in? Are there competitors? Does journalism play a role in these networks, and if so, how? To address this, I gave each interviewee a blank piece of paper and conducted participatory mapping (see Chapter 3). During the process, I ask questions like:

- How would you describe the larger community mySociety is involved in?
- Who do you consider as your peers or colleagues from other organizations?
  - For example, from the Open Knowledge Foundation or other NGOs, or public institutions?
  - Are there international organizations you consider as colleagues or partners?
- Do you see them also as competitors? Or are there any other people or organizations that you see as competitors?
  - What are you competing for?

If journalism was not mentioned, I asked:

- How about journalists. Do you see them as partners or competitors, too? Or neither?
  - Can you explain why, or why not?
  - Are any of the tools you make specifically made for journalists?
  - Are you interested in whether journalists use the tools you create?

7. Debriefing

After thanking my interviewees for their time and turning off the recording device, I asked:

- Is there anything you would like to add? Anything important about mySociety and your work that we did not cover?
- To whom should I talk to next? Can you help me to get in touch with them?

As I was interested in learning some of the techniques they use in their work, I asked my interviewees for advice:

- I think it would really expend my understanding of civic tech when I could learn to do it myself to some extent. Do you have any advice as to what kinds of skills I should learn and where I could pick them up? What skills are important to work at mySociety (such as any types of programming languages that are important)? What are your acquired competences in your work?

Nederlandse samenvatting

Moderne samenlevingen worden, zo wordt vaak gesteld, steeds meer 'gedreven door data'. Dataficatie markeert een omslag van het werken met en vertrouwen op kleine representatieve steekproeven naar een gestage kwantificatie van het sociale leven, waarbij sociale handelingen continu worden omgezet in meetbare data. Of we de veelal revolutionaire retoriek rond dataficatie en big data als vanzelfsprekend beschouwen of niet, commerciële actoren, regeringen en wetenschappers zijn in toenemende mate afhankelijk van het verzamelen en analyseren van data. Hoewel er consensus bestaat over het idee dat dit aanzienlijke gevolgen heeft voor de sociale constellaties die wij vormen, zijn de implicaties ervan voor de democratie, en verschillende groepen daarbinnen, zeer omstreden.

Dit proefschrift onderzoekt hoe data democratische publieken beïnvloedt door te kijken naar de praktijken en *social imaginaries* van twee groepen actoren die het gebruik van kwantitatieve technieken faciliteren op twee kernterreinen in de publieke sfeer: data-activisten en datajournalisten. Data-activisme omvat het activisme in de open data en civic tech-bewegingen dat zich richt op het ontwikkelen van projecten die het makkelijker maken voor burgers om zich te engageren met autoriteiten. Voorbeelden hiervan zijn websites die parlementaire toespraken toegankelijker maken of hulp bieden bij het indienen van verzoeken bij publieke instituties in het kader van vrijheid van informatie. Het label datajournalistiek wordt gebruikt voor het beschrijven van alle vormen van journalistiek waarbij er met kwantitatieve data wordt gewerkt, d.w.z. om onderzoeksreportages te ondersteunen of om interactieve webapplicaties te ontwikkelen die gebruikers in staat stellen om zelf data te verkennen.

Datajournalisten en data-activisten kunnen worden beschouwd als 'pioniersgemeenschappen' die als voorbeeld fungeren voor andere journalisten of actoren uit het maatschappelijk middenveld. Daarnaast hebben datajournalisten en data-activisten een nauwe onderlinge band ontwikkeld, omdat zij beiden afhankelijk zijn van data en een publieke dienst willen leveren. Omdat het waarschijnlijk is dat de door pioniersgemeenschappen ontwikkelde praktijken op grotere schaal zullen worden toegepast, verschaft het bestuderen van de manieren waarop deze actoren elkaar kunnen aanvullen relevante inzichten in hoe data de verhouding tussen de journalistiek en de samenleving in bredere zin kan veranderen. Toch hebben data-activisme en de verhoudingen tussen data-activisten en datajournalisten tot nu toe nog niet veel aandacht gekregen in *media-* en *journalism studies*.

In dit proefschrift onderzoek ik de praktijken en visies van deze actoren, beginnend met een uitgebreide literatuurstudie (Hoofdstuk 2). Ik behandel drie aspecten die relevant zijn voor het analyseren van datajournalisten en data-activisten als aan elkaar gerelateerde pioniersgemeenschappen. Als eerste bespreek ik de historische ontwikkelingen van deze praktijken, waarbij ik uitleg hoe datajournalisten en data-activisten zich tot 'pioniers' hebben ontwikkeld en hoe zij bepaalde tradities voortzetten dan wel doorbreken. Als tweede ga ik in op studies naar de relatie tussen datajournalisten en actoren uit de technologiesector die hun oorsprong vinden in de open source-cultuur (programmeurs, computerwetenschappers, enzovoort). Als derde ga ik na hoe het concept van *monitorial citizenship* zich verhoudt tot deze actoren en bespreek ik de implicaties daarvan voor deze studie.

De methodologische aanpak van deze studie is voornamelijk kwalitatief, multi-methodologisch en geïnspireerd op *practice theory* (Hoofdstuk 3). Om het veld van datajournalistiek en data-activisme in kaart te brengen heb ik etnografisch

werk verricht op conferenties en workshops, en gebruikgemaakt van Digital Methods. De verzameling en analyse van de data is ingegeven door een constructivistische *grounded theory* benadering, en is gebaseerd op interviews, inhoudsanalyse, en participatieve *mapping* om de verwikkelingen tussen datajournalisten en data-activisten grondig te bestuderen. Ik heb drie case studies uitgevoerd (Hoofdstukken 4-6). De eerste twee case studies (Hoofdstukken 4-5) zijn gericht op de non-profitorganisaties die zich bevinden in het veld van data-activisme, de derde (Hoofdstuk 6) vergelijkt data-activisten met van datajournalisten om na te gaan hoe deze groepen met elkaar zijn verweven. In de drie case studies beantwoord ik twee onderzoeksvragen:

Wat is de rol van data in de social imaginaries en praktijken van data-activisten en datajournalisten?

De bevindingen van deze studie tonen aan dat datajournalisten en data-activisten geen nieuwe *social imaginaries* rond data bedenken; hoe zij data zien en gebruiken is in lijn met de twee reeds gevestigde begrippen van de participatieve cultuur en de journalistieke gate-keeping. Wat nieuw is, is dat data wordt ingezet om deze begrippen op innovatieve manieren te implementeren. In Hoofdstukken 4-5 beschrijf ik hoe data-activisten zich bezighouden met het uitdenken en -voeren van wat een 'data-gedreven' vorm van de participatieve cultuur kan worden genoemd, waarin data wordt gezien als een middel ter bevordering van democratische betrokkenheid en wordt ingezet om burgers te faciliteren om autoriteiten te benaderen. Data-activisten willen met behulp van hun tools de drempel voor burgers verlagen om gebruik te maken van overheidsdiensten. Dit moet besluitvormingsprocessen transparanter, participatiever en uiteindelijk representatiever maken. Data-activisten pleiten zelden zelf voor een bepaald beleid; hun doel is om anderen,

zoals maatschappelijke belangengroepen en professionele journalisten, in staat te stellen om dat doeltreffender te doen.

In Hoofdstuk 4 onderzoek ik de *social imaginaries* van activisten die gelieerd zijn aan de Open Knowledge Foundation Germany (OKF DE), en de implicaties van deze verbeeldingen voor de publieken die zij beogen te dienen. Ik laat zien hoe activisten een aantal kernpraktijken van open source (zoals het delen van broncode) overnemen en deze toepassen op het genereren, gebruiken en analyseren van data om zodoende de relatie tussen overheden en hun publieken te veranderen. De OKF DE beschouwt onbewerkte data bijvoorbeeld als een vorm van broncode, waarvan de interpretatie en het hergebruik kennis genereren. Het openlijk delen van onbewerkte data (net zoals de open source beweging softwarecode deelt) zou burgers in staat stellen om op basis van eigen interpretaties van de door overheden verzamelde data beslissingen te nemen, en te rechtvaardigen. Uit mijn bevindingen blijkt hoe open data-activisten alternatieve *social imaginaries* rond dataficatie verwerven en een nieuw begrip ontwikkelen voor gezamenlijke kenniscreatie en -distributie binnen gedataficeerde publieken.

In Hoofdstuk 5 bestudeer ik hoe een andere civic tech-organisatie, de Britse ngo mySociety, data inzet om een visie van een meer participatieve cultuur te realiseren. Ik laat zien dat de manieren waarop leden van mySociety data zien en benutten erop gericht zijn om de agency van publieken ten opzichte van overheden uit te breiden. Het hergebruiken en creëren van data stelt mySociety in staat om tussen overheden en publieken te bemiddelen. Hun diensten moeten fungeren als 'lagen' tussen de burgers en de door overheden gevolgde wettelijke en bureaucratische procedures. In wezen herinterpreteert mySociety hoe bestaande rechtskaders en openbare diensten 'idealiter' zouden moeten functioneren en probeert het invloed uit te oefenen op hoe deze worden

geïmplementeerd. Hun tools en diensten zijn veelal gericht op het scheppen van een soort 'datatransparantie' die de prestaties van overheden op nieuwe manieren inzichtelijker maakt voor het publiek. Hun vrijheid van informatie-websites publiceren bijvoorbeeld alle reacties van autoriteiten op verzoeken om vrijheid van informatie. Zodoende worden de toepassingen van het wettelijk recht op vrijheid van informatie openbaar en zichtbaar. Dit stimuleert autoriteiten om te reageren op de verzoeken van mySociety en helpt burgers om beter te begrijpen hoe ze gebruik kunnen maken van de wet. Daarnaast moet het ervoor zorgen dat de relatie de tussen autoriteiten en het publiek collaboratiever en representatiever wordt.

In Hoofdstuk 6 onderzoek ik de praktijken en verbeeldingen van datajournalisten. Ik laat zien dat de ideeën die zij hebben rond data variëren en afhankelijk zijn van hun opleiding en de organisatorische setting waarin zij werken. We kunnen onderscheid maken tussen grofweg twee belangrijke praktijken. De eerste daarvan is het inzetten van data ter ondersteuning van traditionele journalistieke gate-keeping. Gate-keeping is het selecteren en communiceren van maatschappelijk relevante informatie. De tweede praktijk houdt in dat data wordt ingezet om anderen in staat te stellen actie te ondernemen – een praktijk die weliswaar is ontwikkeld door data-activisten, maar ook zichtbaar is onder datajournalisten. De wijze waarop deze twee praktijken, d.w.z. gate-keeping en het faciliteren van anderen, elkaar kunnen aanvullen, is cruciaal voor het behandelen van de tweede onderzoeksvraag:

Op welke punten convergeren en divergeren de praktijken en social imaginaries van datajournalisten en data-activisten, en hoe bepaalt dit hun onderlinge verwevenheid?

De bevindingen in Hoofdstuk 6 laten zien dat de grenzen tussen data-activisme en datajournalistiek zijn vervaagd: sommige journalisten passen

praktijken en waarden toe die vergelijkbaar zijn met die van data-activisten, en combineren die met journalistieke gate-keeping. Ik beschrijf hoe data-activisten en datajournalisten elkaar zien als complementair, wat soms zorgt voor naadloze samenwerkingen binnen de projecten waar ze zich mee bezighouden. Ik geef voorbeelden van journalistieke onderzoeksprojecten die data-activisten op ideeën brengen voor nieuwe applicaties en situaties waarin data-activisten tools ontwikkelen voor journalisten.

De resultaten van dit onderzoek tonen aan dat datajournalisten en data-activisten een bepaalde praktijkgemeenschap of configuratie vormen, d.w.z. een netwerk van menselijke actoren dat gestoeld is op in elkaar grijpende praktijken en gedeelde waarden. Naast data-activisten identificeer ik drie groepen actoren die werkzaam zijn binnen de professionele journalistiek maar verschillen in de relatieve nadruk die ze leggen op gate-keeping en faciliteren: Normaliseerders, Experimenteerders en Vertalers. Normaliseerders zijn datajournalisten met een formele journalistieke opleiding die werkzaam zijn in gevestigde nationale nieuwsorganisaties. Van de drie genoemde groepen leggen Normaliseerders de meeste nadruk op gate-keeping. Zij zetten data in op manieren die reeds lang bestaande ideeën rond professionele onderzoeksjournalistiek onderschrijven. Data wordt door Normaliseerders gezien en gebruikt als een middel om journalistiek te bedrijven op min of meer dezelfde manier waarop die al tientallen jaren is bedreven, maar dan wel efficiënter en op grotere schaal.

Experimentateerders en Vertalers nemen een afwijkende positie in en combineren gate-keeping met faciliteren. Vertalers zijn actief in zowel de journalistiek als het civic tech-veld en werken binnen nieuwsorganisaties op plekken waar die twee velden elkaar kruisen. Een voorbeeld hiervan dat ik bespreek is de Duitse nieuwsorganisatie Correctiv. Dat is een non-profit

redactie die zowel de professionele onderzoeksjournalistiek wil verbeteren als de burger wil leren om zelf journalist te worden. Enerzijds is Correctiv dus geïnteresseerd in het faciliteren van anderen om eigen onderzoeken te doen, en biedt het hulpmiddelen en trainingen. Tegelijkertijd is hetgeen dat Correctiv wil vergemakkelijken niet het engagement tussen gewone burgers en hun regeringen, maar journalistieke gate-keeping. Gate-keeping en faciliteren moeten elkaar wederzijds versterken. De meeste leden van de laatste groep, die van de Experimenteerders, hebben een achtergrond in de technologiesector en plaatsen vraagtekens bij wat professionele journalistiek inhoudt en waar haar grenzen liggen. Hoewel ze werkzaam zijn in mediaorganisaties, benadrukken ze het faciliteren meer dan journalistieke gate-keeping. Zij willen net als data-activisten hun gebruikers in staat stellen om zelf kwesties te verkennen. Het belangrijkste verschil tussen Experimentateerders en Vertalers is dat de eersten doorgaans niet andere journalisten willen faciliteren in het uitvoeren van hun eigen onderzoeken.

Hoewel 'participatieve cultuur' (die de basis vormt voor faciliterende praktijken) en 'journalistieke gate-keeping' meestal worden opgevat als twee elkaar uitsluitende concepten (de één zorgt voor open participatie, de ander voor professionele controle), laten mijn bevindingen zien dat de praktijken en waarden die aan deze concepten ten grondslag liggen elkaar op innovatieve wijze kunnen aanvullen. De groeiende afhankelijkheid van data zorgt ervoor dat faciliteren en gate-keeping in elkaar grijpen en in een continuüm kunnen bestaan. Data wordt gebruikt om zowel de journalistieke gate-keeping te versterken als gebruikers te faciliteren. Dit heeft belangrijke implicaties voor de theorievorming rond dataficatieprocessen, en voor hoe wij die zouden moeten bestuderen. Op methodologisch vlak stel ik dat we meer aandacht moeten hebben voor hoe dataficatie zich verhoudt tot nieuwe, opkomende

configuraties, zoals die tussen data-activisten en datajournalisten. Op theoretisch vlak moeten we voorzichtig zijn in ons gebruik van reeds bestaande concepten bij het verklaren van transformaties die momenteel nog aan de gang zijn. Nieuwe configuraties stellen bestaande concepten op de proef, en het simpelweg gelijkstellen van data-activisme aan participatieve cultuur, of datajournalistiek aan gate-keeping, gaat voorbij aan het feit dat deze actoren verschillende opvattingen van burgerschap met elkaar combineren, en aan de diverse manieren waarop zij dat doen.

Een toekomstige kritische studie naar dataficatie zal zowel empirisch onderzoek als normatieve theorie vereisen. Empirisch onderzoek naar de praktijken en verbeeldingen van pioniersgemeenschappen, zoals het hier gepresenteerde onderzoek, kan zowel normatieve debatten als de pioniersgemeenschappen zelf van nuttige informatie voorzien. Onderzoekers die geïnteresseerd zijn in normatieve kwesties rond dataficatie en pioniersgemeenschappen hebben vergelijkbare doelen: het behouden of uitbreiden van reeds lang bestaande ideaalbeelden van democratie, gelijkheid en verantwoordelijkheid. Empirisch onderzoek kan helpen bij het ontwikkelen van meer inzicht in de manieren waarop dataficatie van invloed is op de waarden die we willen behouden of uitbreiden, en zodoende een dialoog bewerkstelligen tussen verschillende groepen met overlappende belangen. Ik heb laten zien hoe een dergelijke empirische benadering kan worden ingezet, en hoop daarmee een relevante bijdrage te hebben geleverd aan de huidige discussie rond de implicaties die dataficatie met zich meebrengt voor de samenleving.

Knowing what counts

From news recommendations to smart cities, our lives are increasingly affected by the aggregation of data. In this dissertation, I study how this growing reliance on data affects democratic visions and practices by looking at the practices, self-understandings, and visions of data journalists and data activists. These actors are particularly relevant because they are pioneers for how data is employed in key areas of democratic publics: journalism and civil society. Data journalists and data activists act as exemplars for other journalists and civil society actors, and thereby shape their perception and use of data. In addition, data journalists and data activists have developed a close relationship with each other because both rely on data, and aspire to provide a public service that empowers citizens. I examine how they are able to collaborate and complement each other, and reflect on the broader implications of their entanglements for the relationship between journalism and civil society.

The dissertation is based on three case studies. The first study shows how data activists draw on open source culture, and envision that freely available government data increases transparency and makes governments more participatory and representative. The second study examines how data activists attempt to enable the public to more easily engage with authorities, using data. The third study shows how the practices of data journalists and data activists interlock and range from the aim to steer public debates by highlighting important issues via gatekeeping on the one hand, to facilitating the actions of citizens on the other. Throughout these three case studies, I show that the growing reliance of data in journalism and civil society affects both how, and by whom, older democratic visions and notions of journalism are being implemented.

ISBN: 978-94-034-0572-8